PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12Q 1/68, C12N 15/12, C07K 14/47, A61K 38/43, C12N 9/00, A01K 67/027, A61K 48/00, C12N 5/10

A2

(11) International Publication Number: WO 97/12992

(43) International Publication Date: 10 April 1997 (10.04.97)

(21) International Application Number: PCT/IB96/01106

(22) International Filing Date: 2 October 1996 (02.10.96)

(30) Priority Data:

9520080.4	2 October 1995 (02.10.95)	GB
60/009,832	11 January 1996 (11.01.96)	US

(71) Applicants: ROYAL NETHERLANDS ACADEMY OF ARTS AND SCIENCES [NL/NL]; Kloveniersburgwal 29, NL-1011 JV Amsterdam (NL). STATE UNIVERSITY OF UTRECHT [NL/NL]; P.O. Box 80125, NL-3508 TC Utrecht (NL). ERASMUS UNIVERSITY ROTTERDAM [NL/NL]; P.O. Box 1738, NL-3000 DR Rotterdam (NL).

(72) Inventors: VAN LEEUWEN, Frederik, Willem; Netherlands Institute for Brain Research, Meibergdreef 33, NL-1105 AZ Amsterdam (NL). BURBACH, Johannes, Peter, Henri; Rudolf Magnus Institute for Neurosciences, University of Utrecht, Universiteitsweg 100, NL-3584 CG Utrecht (NL). GROSVELD, Franklin, Gerardus; Dept. of Cell Biology and Genetics, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam (NL).

(74) Agent: HALLYBONE, Huw, George; Carpmaels & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: DIAGNOSIS METHOD AND REAGENTS

(57) Abstract

The invention encompasses methods and reagents for the diagnosis of a disease caused by or associated with a gene having a somatic mutation giving rise to a frameshift mutation. The methods include the steps of providing a body fluid or tissue sample from a patient; and analyzing the sample for the presence of a gene having a frameshift mutation or a protein encoded thereby, wherein the presence of the mutated gene or encoded protein is indicative of the disease.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia
AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CS	Czechoslovakia
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakhstan
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam

DIAGNOSIS METHOD AND REAGENTS

The invention encompasses methods and reagents for the diagnosis of a disease caused by or associated with a gene having a somatic mutation giving rise to a frameshift mutation. The methods include the steps of providing a body fluid or tissue sample from a patient; and analyzing the sample for the presence of a gene having a frameshift mutation or a protein encoded thereby, wherein the presence of the mutated gene or encoded protein is indicative of the disease.

BACKGROUND OF THE INVENTION

Many diseases are believed to stem from somatic genetic mutation rather than inherited gene abnormalities, including different types of cancers and neurodegenerative diseases.

Diseases which are caused by somatic mutation are often age-related, with an increasing incidence of somatic mutation with increasing age. For example, the prevalence of cancer increases with age, and it appears that some cancers are caused by somatic mutation. Cancer is the second leading cause of death in the United States, accounting for approximately 500,000 deaths (or 20 percent of all deaths) per year. These diseases are so prevalent that unless current trends are reversed, one in three living Americans will develop cancer at some time. Cancers are usually detected by clinical methods and cytological methods and are difficult to detect early. In some cases, familial cancers have been traced to inherited genes, and detection of such genes is believed to be predictive of susceptibility to the cancer.

It is an object of the present invention to provide methods and assays for detection and/or treatment of diseases involving genetic mutations, particularly those diseases relating to aging, wherein the probability of having the disease increases with the age of the patient. The invention contemplates detection and/or treatment of those age related diseases which are due to mutations occurring in the DNA of somatic cells. If the mutations are not corrected, the disease may result.

Another object of the invention is to treat diseases identified according to the invention, by providing to a patient afflicted with the disease or having a propensity to develop the disease, a corrective agent such as an enzyme or oligonucleotide.

Yet another object of the invention is to provide a method for identifying age-related diseases by correlating nucleotide sequence mutation hotspots with the disease.

Other objects of the invention relate to identification, detection and treatment of age related diseases including cancers (especially non-hereditary cancers) and neurodegenerative diseases, such as Alzhemier's Disease (AD), Parkinson's Disease (PD), Down's syndrome, frontal lobe dementia (Pick's Disease), progressive supranuclear palsy (PSP), amyotrophic lateral sclerosis, Huntington's Disease, multiple sclerosis, and other degenerative diseases. Early disease diagnosis is important for effective treatment.

Alzheimer's Disease is in most cases a disease which is related to aging. AD is characterised by atrophy of nerve cells in the cerebral cortex, subcortical areas, and hippocampus and the presence of plaques, dystrophic neurites and neurofibrillary tangles. In most cases, it is not known whether AD is caused by a genetically inherited trait or by environmental factors, e.g., somatic mutations, or both.

The pathogenic mutation is unknown.

Another object of the invention is to provide a diagnostic test for AD which enables definitive diagnosis of AD in living patients. Furthermore, as AD is a progressive disease, it is desirable to diagnose AD as early as possible so that preventative action may be taken.

A number of diagnostic methods have been previously suggested for AD diagnosis, most of which have focused on the amyloid protein. See for example U.S. Patents 4,666,829, 4,816,416 and 4,933,159. However, amyloid deposits have been found in individuals, especially aged persons, who have not shown signs of dementia (See J. Biol. Chem., 265, pp 15977, 1990; and Tables 2 and 3). Diagnostic tests based on the amyloid protein have therefore been shown to lack specificity for AD.

In U.S. Patent 4,727,041 a diagnostic test for AD is described based on measuring levels of somatotropin and somatomedin-C in blood sera following administration of an L-dopa proactive test.

In International patent application WO 94/02851, a method is described for identifying AD by the use of antibodies having affinity for paired helical filaments in order to determine the levels of paired helical filaments in cerebral spinal fluid. The presence of paired helical filaments is alleged to be indicative of AD.

Other diagnostic methods are based on the identification of "disease specific marker proteins" in the cerebral spinal fluid. In International patent application WO 95/05604, for example, five disease specific proteins are identified by their molecular weights. However, the specific identity of the proteins is unknown and their specific relationship to the pathogenesis of AD is also unknown. The five "disease specific marker proteins" may

therefore be present as a result of a more fundamental cellular or biochemical change.

Another object of the invention is to provide for detection of AD preferably early on in the disease state. It is desirable to detect a protein or substance which is either directly responsible for the disease or is involved early on in the pathogenesis of the disease, or if not involved is nevertheless generated directly or indirectly by the mechanism causing the disease. Such a protein or substance may be the "causative" agent to the disease or may be "associated with" the disease state in the sense of being diagnostic of the disease state.

Recently, Sherrington *et al.* in Nature, 375, pp 254-260, 1995, identified a gene on chromosome 14 bearing missense mutations which are associated with up to 70% of familial early onset AD cases. A missense mutation involves a nucleotide substitution, usually a single nucleotide substitution, which results in an amino acid substitution at the corresponding codon. The missense mutations disclosed in Sherrington *et al.* are predicted to change the encoded amino acid at the following positions (numbering from the first putative initiation codon) Met to Leu at codon 146, His to Arg at codon 163, Ala to Glu at codon 246, Leu to Val at codon 286, Cys to Tyr at codon 410. It has been proposed that these mutations may be useful in identifying early onset AD. As stated earlier, the majority of AD cases are late onset (after 65 years of age; Table 1) and it is therefore still a problem to identify the majority of individuals having AD, particularly late onset AD.

Presently, there are a number of substances which are alleged to be useful in the treatment of AD. However, so far only limited success has been achieved with these substances. Methods for effectively treating and/or preventing AD are still required (see Allen and Burns, Journal of Psychopharmacology, 9, pp 43-56, 1995).

SUMMARY OF THE INVENTION

The present invention is based on the observation that a gene containing a frameshift mutation and encoding a corresponding mutant protein may be correlated with the presence of a disease.

According to the present invention there is provided a method for the diagnosis of a disease caused by or associated with at least one gene having one or more somatic mutations giving rise to a frameshift mutation comprising: i. providing a biological sample, such as a body fluid or tissue sample, from a patient; and ii. analyzing the sample for the presence of a gene having a frameshift mutation or a mutant protein encoded thereby, wherein the presence of the mutated gene or mutant protein is indicative of the disease.

The term "somatic mutation" refers to a mutation occurring in a somatic or non-germline cell, and does not pass through the germline and is therefore not inherited.

A "mutant" protein is a polypeptide encoded by a mRNA at least a part of which is in a reading frame that is shifted relative to the initiation start codon from that of the native or wild-type reading frame, and thus will include any protein having an aberrant carboxy terminal portion which is encoded by the +1 or +2 reading frame of the wild type gene sequence. Thus, the mutant protein will include a hybrid wild-type/nonsense protein having an amino terminal amino acid sequence that is encoded by the wild type (0) reading frame and a carboxy terminal amino acid sequence that is encoded by the +1 or +2 reading frame, and thus the nonsense portion of the mutant protein. The cross-over point between the wild type and nonsense amino acid sequences is the codon containing the frameshift mutation.

The invention is based on the discovery of the presence of such a mutant protein or an accumulation of more than one mutant protein in a tissue from a diseased individual, and also on identification of the mutant protein as indicative of the disease.

The phrase "caused by or associated with" refers to a gene which is either fully or partly responsible for the disease, or a gene which is not responsible for the disease but is associated with the diseased state in the sense that it is diagnostic of the diseased state.

A disease caused by or associated with at least one gene having one or more somatic mutations giving rise to a frameshift mutation can be any disease including non-hereditary cancers, neurodegenerative diseases such as Parkinson's Disease (PD), Alzheimer's Disease (AD), Down's syndrome, frontal lobe dementia (Pick's Disease), progressive supranuclear palsy (PSP), amyotrophic lateral sclerosis, Huntington's Disease, multiple sclerosis, and other degenerative diseases such as cardiovascular diseases and rheumatoid arthritis. Cancers treatable according to the invention include but are not limited to Hodgkin's disease, acute and chronic lymphocytic leukemias, multiple myeloma, breast, ovary, lung, and stomach or bladder cancers.

A gene having a somatic mutation which leads to a frameshift mutation, and herein referred to as the "mutant gene", can be any gene having at least one mutation which leads to a frameshift mutation.

A "frameshift mutation" refers to a deletion or insertion of one or more nucleotides within an open reading frame, for example, a single nucleotide or dinucleotide deletion or insertion, such that the reading frame of the coding region is shifted by one or two nucleotides.

Preferably, the frameshift mutation is a nucleotide or dinucleotide deletion leading to a + 1 or +2 frameshift mutation. However, any number of nucleotide deletions can occur provided a frameshift mutation results. Alternatively, the insertion of one or more nucleotides may give rise to a frameshift and such mutations also form part of the present invention.

Other genetic modifications which give rise to a frameshift also form part of the present invention, such as a change in the nucleotide sequence which leads to translation initiation from a different position or a mutation outside a coding region, such as within an Intron or a 5' or 3' untranslated region, which mutation may result in mis-translation and production of a mutant protein. In this type of gene mutation, the mutant protein would be completely nonsense sequences and would contain no wild-type sequences.

It is preferred that the mutation is a nucleotide and more preferably a dinucleotide deletion or insertion associated with the nucleotide sequence GAGA of the gene; especially preferred frameshift mutations are associated with the nucleotide sequence GAGAX, where X is one of G, A, T or C; thus preferred motifs include GAGAG, GAGAC, GAGAA, and GAGAT. Preferably the dinucleotide deletion is an AG deletion. It is further preferred that the mutant gene has one or two dinucleotide deletions associated with a GAGA, GAGAG, GAGAC, GAGAA, or GAGAT leading to a + 1 or + 2 frameshift mutation respectively.

In a preferred embodiment of the invention, the somatic mutations occur in genes of the neuronal system, where the disease is a neurodegenerative disease.

The "neuronal system" is defined as any cells, genes, proteins or substances relating to or forming part of the neuronal system such as nerve cells, glial cells, proteins

including Tau, ß amyloid precursor protein, ubiquitin, apolipoprotein E4 neurofilament proteins and microtubule associated protein II, and the genes encoding the proteins.

Where the disease is a neurodegenerative disease, especially AD, the preferred mutant genes of the present invention are those encoding the ß amyloid precursor protein, the Tau protein, ubiquitin, apolipoprotein-E_4 (Apo-E_4), microtubule associated protein II (MAP 2) and the neurofilament proteins, having a deletion, insertion or other modification leading to a frameshift mutation. The most preferred mutant genes of the present invention are those encoding ubiquitin, MAP 2 and the neurofilament proteins, having a frameshift mutation.

It is preferred that the mutation is an AG dinucleotide deletion associated with a GAGA, or GAGAX leading to a frameshift mutation. It is further preferred that the mutant gene has one or two AG deletions each associated with a GAGA or similar motif, leading to a + 1 or + 2 frameshift mutation respectively.

The term "mutant protein" as used herein is defined as the protein encoded by the mutant gene of the present invention.

It is preferred that the methods of the present invention are for the diagnosis of a disease caused by or associated with at least one gene having one or more somatic mutations giving rise to a frameshift mutation. A preferred disease for diagnosis by the present invention is AD, except the early onset AD cases found to be linked to chromosome 14 and 1. It is further preferred that the methods of the present invention are for the diagnosis of late onset AD, especially non-familial or "sporadic" late onset AD cases.

As used herein, "biological sample" refers to a body fluid or body tissue which contains proteins and/or cells

from which nucleic acids and proteins can be isolated. Preferred sources include buccal swabs, blood, sperm, epithelial or other tissue, milk, urine, cerebrospinal fluid, sputum, fecal matter, lung aspirates, throat swabs, genital swabs and exudates, rectal swabs, and nasopharyngeal aspirates.

The body fluid sample can be any body fluid which contains cells having the somatic mutation which gives rise to the frameshift mutation and causes the diseases. When the disease is a neurodegenerative disease it is preferred that the body fluid sample contains cells of the neuronal system or the products of such cells. When the disease is a neurodegenerative disease, the preferred body fluid is cerebral spinal fluid which can be obtained after a lumbar puncture (Lannfelt *et al.*, Nature Medicine, 1, pp 829-832, 1995). Another preferred body fluid is blood, as it is easily obtained and contains lymphocytes which can be analyzed for the presence of the mutant gene or encoded protein.

The tissue sample can be any tissue and is preferably one that can be easily obtained, such as skin and nose epithelium.

Preferably, when analyzing the sample for a mutant gene, a nucleic acid probe is used. The nucleic acid probe is preferably a nucleotide probe having a sequence complementary to part of the mutant gene encompassing the mutation giving rise to the frameshift mutation.

As would be apparent to one skilled in the art, the probe may be used to detect DNA or RNA in a fluid sample or in a tissue sample.

The present invention further provides a nucleic acid probe having a sequence complementary to part of the mutant gene encompassing the mutation leading to the frameshift

mutation. The probe is preferably sufficiently complementary to the mutant sequence of the gene so that under stringent conditions the probe only remains bound to the mutant sequence. "Stringent" conditions are defined herein as DNA:DNA hybridization conditions which may be performed at 65°C using a hybridization buffer equivalent to 50% formamide and 0.1X SSC (see below and Evans et al. PNAS (1994) 9; 6059-6063, 6060). "Stringent" conditions also preferably include stringent washes, as described in Evans et al. (Ibid).

The probe may be of any length but is preferably between 5 and 50 nucleotides long, more preferably between 10 and 30 nucleotides long. For example, the probe may be 5, 10, 15, 20, 25, or 30 nucleotides in length.

In a preferred embodiment the probe comprises a sequence complementary to a GAGA or GAGAX, having a nucleotide or dinucleotide deletion or insertion, and nucleotide sequences corresponding to the nucleotide sequences flanking the GAGA or motif in the wild-type gene. It would be apparent to one skilled in the art that if RNA was being probed for, a probe comprising a sequence complementary to the corresponding GAGA motif present in the RNA would have to be used.

Methods of detecting the presence of the mutant gene include the polymerase chain reaction (PCR) using primers having a sequence complementary to the sequence either side of the mutation which gives rise to the frameshift mutation are used to amplify the DNA, as described hereinbelow.

The primers used in the above PCR based method can vary in size from 20bp to 2-3 kb; for example, 20bp, 50bp, 100bp, 500bp, 1000bp, 1500bp, 2000bp, or 3000bp. The primers can be prepared by a number of standard techniques including cloning the sequences flanking the nucleotide region to be amplified or by synthesizing the primers using

phosphoramidite method.

The present invention further provides primers for use in the above defined PCR based methods for the amplification of the nucleotide region containing the mutation.

Preferably, when analyzing the sample for the mutant protein of the present invention an immunological test is employed. The immunological test is preferably based on the use an antibody molecule having specificity for the mutant protein of the present invention and not the wild-typ protein.

The present invention thus further provides an antibody molecule having specificity for the mutated protein of the present invention but not for the wild-type protein. Prefereably, the antibody is specific for the carboxy terminal end of the mutant protein.

The present invention further provides a method for the diagnosis of a neurodegenerative disease comprising: i. providing a body fluid or tissue sample from a patient; and ii. analyzing the sample for the presence of a gene of the neuronal system having a frameshift mutation or a protein encoded thereby, wherein the presence of the mutated gene is indicative of a neurodegenerative disease.

Preferably, the neurodegenerative disease is AD.

The present invention also relates to methods for preventing and/or treating the diseases, vectors for preventing and/or treating the diseases and for the production of diagnostic reagents, compositions for preventing and/or treating the diseases, nucleic acid sequences, probes and antibody molecules for use in the present invention and transgenic animals.

The present invention further provides a diagnostic kit for diagnosing a disease caused by or associated with at least one gene having one or more somatic mutations giving rise to a frameshift mutation comprising: i. a nucleic acid probe having a sequence complementary to part of the mutant gene which encompasses the mutation which leads to the frameshift mutation and packaging materials therefor; and ii. means for detecting the probe bound to the mutant gene.

The present invention further provides a diagnostic kit for diagnosing a disease caused by or associated with at least one gene having one or more somatic mutations giving rise to a frameshift mutation comprising: i. primers for use in a PCR reaction, the primers having a sequence complementary to the sequence either side of the mutation which gives rise to the frameshift mutation, packaging materials therefor, and reagents necessary for performing a PCR reaction and amplifying the DNA or RNA sequence containing the mutation; and ii. means for detecting the amplified DNA or RNA sequence containing the mutation.

The present invention further provides a diagnostic kit for diagnosing a disease caused by or associated with at least one gene having one or more somatic mutations giving rise to a frameshift mutation comprising: i. an antibody molecule having specificity for the mutant protein of the present invention and not the wild-type protein; and ii. means for detecting the antibody molecule bound to the mutant protein.

The antibody molecule and the means for detecting the bound antibody molecule are as defined above.

In a further embodiment of the present invention the diagnostic kit described above additionally comprising: i. an antibody molecule having specificity for the wild-type protein; and ii. means for detecting the antibody molecule bound to the wild-type protein, as a control for diagnosing

a disease caused by or associated with at least one gene having one or more somatic mutations giving rise to a frameshift mutation.

The present invention further provides a gene having one or more somatic mutations giving rise to a frameshift mutation which causes or is associated with a disease.

The present invention further provides a mutated protein encoded by the mutated gene found to be indicative of a disease , the mutant gene having one or more somatic mutations giving rise to a frameshift mutation. Preferably, the mutant protein contains an antigenic epitope specific for the diseased state, examples of which are provided in Table 7.

In a preferred embodiment of the present invention the mutated gene encodes a protein comprising at least part of the sequence designated +1 or +2 in any one of Figures 2 to 9, or an immunologically equivalent fragment thereof.

In a preferred embodiment the mutated protein comprises any one of the following individual sequences: RGRTSSKELA; HGRLAPARHAS; YADLREDPDRQ; RQDHHPGSGAQ; GAPRLPPAQAA; KTRFQRKGPS; PGNRSMGHE; EAEGGSRS; or VGAARDSRAA, (Seq. ID Nos: 1-9, respectively) especially when the disease is a neurodegenerative disease such as AD.

In a preferred embodiment, the antibody molecule of the present invention has affinity for the mutant proteins defined above.

The present invention also relates to a method for treating and/or preventing a disease caused by or associated with at least one gene having one or more somatic mutations giving rise to a frameshift mutation. The finding of mutations in genes which lead to the production of mutant proteins, and which are indicative of a disease, has led to

a number of ways of treating and/or preventing the disease.

The present invention further provides a method for detecting errors in nucleic acid repair mechanisms comprising: i. providing a body fluid or tissue sample from a patient; and ii. analyzing the sample for the presence of a gene having a frameshift mutation or a protein encoded thereby, the presence of a mutation being indicative of an error in the nucleic acid repair mechanisms.

The correction of the mutations found in the mutant genes of the present invention is therefore a valuable method for combatting diseases.

The present invention further provides methods of treating a disease by administering to a patient an expression vector encoding one or more DNA repair enzymes.

The present invention further provides the use of an expression vector encoding one or more repair enzymes in the manufacture of a composition for the treatment of a disease caused by or associated with at least one gene having one or more somatic mutations giving rise to a frameshift mutation.

The present invention further provides a method of treatment and/or prevention of a disease caused by or associated with at least one gene having one or more somatic mutations giving rise to a frameshift mutation, comprising administering an expression vector encoding one or more repair enzymes to a patient suffering from or likely to suffer from the disease.

The present invention further provides an expression vector encoding one or more repair enzymes.

The repair enzymes encoded by the vector can be any repair enzyme and/or an associated protein (see Kunkel, Current Biology, 1995, Vol. 5, No. 10, p.1091-1094); such

as, for example, a G/T mismatch binding protein (GTBP or p160), which are capable of or contribute to the repair of a somatic mutation which leads to a frameshift mutation, such as a dinucleotide deletion. Preferred repair enzymes include Mut H, Mut S, Mut L, and Mut U, and the homologs thereof, including mammalian homologs. Such homologs include MSH 1-6, PMS 1-2, MLH 1 and GTBP.

The invention also encompasses methods of combatting diseases caused by at least one gene having one or more somatic mutations giving rise to a frameshift mutation by targeting the mRNA transcript transcribed from the mutat d gene, and correcting the mutant mRNA using ribozymes.

Accordingly, the present invention further provides methods of treating the disease by administering an expression vector encoding a ribozyme in therapy.

The present invention further provides the use of an expression vector encoding a ribozyme in the manufacture of a composition for the treatment of a disease caused by at least one gene having one or more somatic mutations giving rise to a frameshift mutation.

The present invention further provides a method of treatment and/or prevention of a disease caused by at least one gene having one or more somatic mutations giving rise to a frameshift mutation, comprising administering an expression vector encoding a ribozyme under the control of a promoter to a patient suffering from or likely to suffer from the disease.

The present invention further provides an expression vector encoding a ribozyme under the control of a promoter.

The ribozyme encoded by the vector is preferably specific for RNA containing the mutations described above or for the RNA encoding a defective repair enzyme. For

example, if a defect in a repair enzyme has lead to the mutated gene, by correcting the transcribed message from the repair enzyme gene the function of the repair enzyme can be restored.

A further method for combatting a disease caused by or associated with at least one gene having one or more somatic mutations giving rise to a frameshift mutation, is by gene therapy.

A vector encoding a non-mutated version of the mutated gene under the control of a promoter and other selected transcriptional or translational control elements can be delivered to affected or susceptible cells leading to the production of the correct protein in the cell. By increasing the percentage of the correct protein produced in relation to the mutated protein the further progression of the disease will be reduced, prevented and possibly reversed.

The present invention further provides methods of treating a disease by administering an expression vector encoding the wild-type version of the mutated gene.

The present invention further provides the use of an expression vector encoding the wild-type version of the mutated gene in the manufacture of a composition for treating a disease caused by or associated with at least one gene having one or more somatic mutations giving rise to a frameshift mutation.

The present invention further provides a method of treatment and/or prevention of a disease caused by or associated with at least one gene having one or more somatic mutations giving rise to a frameshift mutation, comprising administering an expression vector encoding the wild-type version of the mutated gene to a patient suffering from or likely to suffer from the disease.

The present invention further provides an expression vector encoding the wild-type version of the mutated gene.

The present invention also provides a vector encoding a wild-type version of the defective repair enzyme which has lead to or contributed to the presence of the mutated gene which causes or is associated with a disease, under the control of a promoter.

Preferably the wild-type repair enzyme is MSH 1-6, PMS 1-2 or MLH 1, or an associated protein such as GTBP.

The present invention further provides the use of more than one of the vectors of the present invention in any combination in therapy.

The present invention further provides the use of more than one of the vectors of the present invention in any combination in the treatment and/or prevention of a disease caused by or associated with at least one gene having one or more somatic mutations giving rise to a frameshift mutation.

The pharmaceutical compositions of the present invention in addition to the vectors as defined herein may additionally comprise a pharmaceutically acceptable excipient.

The invention also encompasses host cell lines and transgenic animals wherein the somatic cells of the animal and the DNA of the host cell contains a transgene which is a mutant gene having one or more somatic mutations giving rise to a frameshift mutation.

Preferably, the transgenic animals of the present invention are additionally defective in nucleic acid repair mechanisms. Thus, the mutated gene specifically expressed in this type of transgenic animal will not be repaired.

The present invention further provides an animal having somatic cells containing at least one of the mutated genes of the present invention.

Preferably, the animal is a rodent, more preferably a rat or mouse.

As used herein, and "expression vector" or an "expressible gene" denotes a vector containing a gene, or a gene which is expressible in a selected host cell. The gene will therefore be operatively associated with transcriptional and translational control sequences sufficient to permit expression of the gene in the host cell. Such control sequences include but are not limited to a promoter, and enhancer, a locus control region, a ribosome binding site, a polyadenylation site and a transcription termination site.

The present invention further provides a method for identifying diseases caused by or associated with at least one gene having one or more somatic mutations giving rise to a frameshift mutation. The method comprises: i. providing the sequence of a gene suspected of being involved in the pathogenesis of a disease; ii. identifying the sequence of the mutant protein encoded by the gene sequence following a frameshift mutation; iii. preparing a probe to the mutant protein or a fragment thereof; and iv. probing a body fluid or tissue sample from a patient having the disease and a patient not having the disease, in order to find a correlation between the presence of the mutant protein and the diseased state.

Preferably, the probe is an antibody molecule as defined herein. It is further preferred that the antibody molecule has affinity for a protein comprising at least one of the sequences: RGRTSSKELA; HGRLAPARHAS; YADLREDPDRQ; RQDHHPGSGAQ; GAPRLPPAQAA; KTRFQRKGPS; PGNRSMGHE; EAEGGSRS;

or VGAARDSRAA, especially when the disease is a neurodegenerative disease such as AD.

Other features and advantages of the invention will be apparent from the following description of the preferred embodiments thereof, and from the claims.

DESCRIPTION OF DRAWINGS

The invention is now illustrated in the appended example with reference to the following drawings:

Figure 1 is a copy of a paraffin section of the frontal cortex of a female Alzheimer patient (70 years old, #83170, Table 2) immunocytochemically incubated with an antibody against a peptide predicted by the + 1 reading frame of βAPP (Tables 5 and 6). Dystrophic neurites (arrowheads) and tangles (arrows) are clearly visible in the cortical layer III.

Figure 2 presents the coding nucleotide sequence of the human amyloid A4 gene (SEQ ID No. 10), the amino acid sequence of the wild-type protein (SEQ ID No. 11), the mutant + 1 frameshift protein (SEQ ID No. 12) and the mutant + 2 frameshift protein (SEQ ID No. 13).

Figure 3 presents the coding nucleotide sequence of the human microtubule-associated protein tau gene (SEQ ID No. 14), the amino acid sequence of the wild-type protein (SEQ ID No. 15), the mutant + 1 frameshift protein (SEQ ID No. 16) and the mutant + 2 frameshift protein (SEQ ID No. 17).

Figure 4 presents the coding nucleotide sequence of the human ubiquitin gene (SEQ ID No. 18), the amino acid sequence of the wild-type protein (SEQ ID No. 19), the mutant + 1 frameshift protein (SEQ ID No. 20) and the mutant + 2 frameshift protein (SEQ ID No. 21).

Figure 5 presents the coding nucleotide sequence of the human apolipoprotein E gene (SEQ ID No. 22), the amino

acid sequence of the wild-type protein (SEQ ID No. 23), the mutant + 1 frameshift protein (SEQ ID No. 24) and the mutant + 2 frameshift protein (SEQ ID No. 25). Information concerning restriction enzyme sites is also given.

Figure 6 presents the coding nucleotide sequence of the human microtubule-associated protein 2 (SEQ ID No. 26), the amino acid sequence of the wild-type protein (SEQ ID No. 27), the mutant + 1 frameshift protein (SEQ ID No. 28) and the mutant + 2 frameshift protein (SEQ ID No. 29).

Figure 7 presents the coding nucleotide sequence of the human neurofilament subunit NF-low (SEQ ID No. 30), the amino acid sequence of the wild-type protein (SEQ ID No. 31), the mutant + 1 frameshift protein (SEQ ID No. 32) and the mutant + 2 frameshift protein (SEQ ID No. 33).

Figure 8 presents the coding nucleotide sequence of the human neurofilament subunit NF-M (SEQ ID No. 34), the amino acid sequence of the wild-type protein (SEQ ID No. 35), the mutant + 1 frameshift protein (SEQ ID No. 36) and the mutant + 2 frameshift protein (SEQ ID No. 37).

Figure 9 presents the coding nucleotide sequence of the human neurofilament subunit NF-H (SEQ ID No. 38), the amino acid sequence of the wild-type protein (SEQ ID No. 39), the mutant + 1 frameshift protein (SEQ ID No. 40) and the mutant + 2 frameshift protein (SEQ ID No. 41).

Figure 10 presents the partial mRNA nucleotide sequence and amino acid sequence of three human neuronal proteins (ß amyloid precursor protein (exons 9 and 10), Tau (exon 13) and Ubiquitin B (exon 2)) expressed in the wildtype and +1 reading frame.

DESCRIPTION

The invention is illustrated by the following nonlimiting examples wherein the following materials and methods are employed. The entire disclosure of each of the literature references cited hereinafter are incorporated by reference herein.

The present invention is based on the discovery that frameshift mutations occur in a single gene or number of genes whose product or products are mutant proteins that are associated with, and indicative of, a disease state. The invention is based on the recognition that the presence of a frameshift mutation results in a new coding sequence for the cell containing the frameshift mutation, and thus a new polypeptide (herein termed a mutant protein) which may be correlated with and thus be indicative of a disease.

According to the present invention, diagnosis and/or identification of a disease caused by or associated with at least one gene having one or more somatic mutations which give rise to a frameshift mutation is accomplished as described herein.

According to the present invention, methods for preventing and/or treating the diseases, vectors for preventing and/or treating the diseases and for the production of diagnostic reagents, compositions for preventing and/or treating the diseases, nucleic acid sequences, probes and antibody molecules for use in the present invention and transgenic animals are accomplished as described herein.

Nucleotide deletions and other somatic mutations occur in genes due to a variety of reasons including errors in replication, as a result of recombination events, due to the presence of the mutagenic compounds, or as a result of a highly transcriptionally active state. Somatic mutations therefore regularly occur. In order to protect against such somatic mutations, mechanisms for correcting such mutations exist. The correction mechanisms involve a number of repair enzymes which detect and bind to the site of a mutation, and correct the mutation.

If the repair enzymes are defective or absent a number of problems can occur. For example, in xeroderma pigmentosum a deficiency in the excision-repair of somatic mutations leads to skin disorders in the patient. More recently, it has been shown that defective repair enzymes can lead to predisposition for certain sporadic cancers. For example, hereditary non-polyposis colorectal carcinoma (HNPCC). See TIG, 10 (5), pp 164 to 168, (1994), for a review.

According to the present invention, methods for detecting errors in nucleic acid repair mechanisms are accomplished as described herein. The correction of the mutations found in the mutant genes of the present invention is therefore a valuable method for combatting diseases.

Further methods for combatting diseases caused by at least one gene having one or more somatic mutations giving rise to a frameshift mutation, include targeting the mutant gene with an oligonucleotide having a correct nucleotide sequence and relying on endogenous or exogenous cellular repair enzymes to repair enzymes to repair the mutation, or targeting the mRNA transcript transcribed from the mutated gene, and correcting the mutant mRNA using ribozymes.

Methods and reagents for disease diagnosis and treatment are described in more detail hereinbelow.

Diagnosis of Diseases According to the Invention

The invention relates to methods for diagnosing diseases caused by or associated with at least one gene having one or more somatic mutations which give rise to a frameshift mutation. Such diseases include but are not limited to cancers and neurodegenerative diseases such as Parkinson's Disease (PD), Alzheimer's Disease (AD), frontal lobe dementia (Pick's Disease), progressive supranuclear

palsy (PSP), amyotrophic lateral sclerosis, Huntington's Disease, multiple sclerosis, Down's syndrome, and other degenerative diseases such as cardiovascular diseases and rheumatoid arthritis.

Somatic mutations can result in a different gene function and have been implicated in diseases associated with ageing, such as certain cancers. However, it has generally been assumed that non-proliferating cells do not undergo important changes at the genomic level. For example, it was assumed previously that genomic changes are mainly related to cell proliferation (Smith, Mutation Research, 277, pp 139-142, 1992) which for non-proliferating cells such as most neurons ends during early postnatal life (Rakic, Science, 277, pp 1054-1056, 1985). However, Evans et al., 1994, Proc. Nat. Aca. Sci. 91:6059, observed that somatic mutations do occur in genes of the neuronal system, i.e., in post-mitotic neurons. The di/di Brattleboro rat, which suffers from severe diabetes insipidus due to the absence of the antidiuretic hormone vasopressin (VP), was the subject of the Evans et al. paper. It had previously been established that the VP hormone was absent in the Brattleboro rat due to a deletion of a single G residue in the second exon of the VP gene, resulting in a mutant VP precursor with an altered C-terminal amino acid sequence. It had also been observed that a small number of neurons in the di/di rat exhibited a heterozygous +/di phenotype and expressed an apparently normal VP gene product. In studying the molecular biology of the di/di rat, Evans et al. identified sequence alterations that restored the reading frame of the mutant VP precursor mRNA, which were based on a two-nucleotide deletion in a GAGAG motif. They correlated the presence of small amounts of normal VP gene product in single magnocellular neurons with a reversion of the mutant gene stemming from a frameshift mutation. Evans et al. concluded that, because +1 frameshift mutations are present in VP transcripts of both wild-type rats and di/di rats, the events leading to these mutations are not caused by the

diseased state of the di/di rat per se. Thus, Evans et al. did not correlate a mutational GAGAG hotspot with a disease state, or predeliction to a disease.

In the present invention, the observations of Evans et al., as to reversions at GAGAG hotspots in VP transcripts within single neurons of the di/di rat leading to wild-type VP gene products, is extended and developed. According to the present invention, a human disease which is caused by or associated with at least one gene having one or more somatic mutations occurring at a mutational hotspot and which give rise to a frameshift mutation is identified and/or diagnosed. The nucleotide sequence of a gene suspected of being involved in the pathogenesis of a disease is provided, e.g., from published gene sequences or from cloning and sequencing of a suspect gene. The amino acid sequence encoded by the gene is then predicted, as are amino acid sequences of encoded mutant proteins. Mutant protein sequences are predicted in +1 and +2 reading frames following a hypothesized frameshift mutation. The location of the frameshift mutation may be hypothesized with respect to certain nucleotide sequence motifs which are suspected of causing frameshift mutations during transcription, examples of such motifs including but not limited to GAGA, for example, GAGAG, GAGAC, GAGAA, and GAGAT.

A probe is then prepared that is specific for the mutant protein or an immunogenic fragment thereof (such probes being described hereinabove for detection of proteins or protein fragments). Depending on where the mutation that leads to the frameshift occurs, part of the mutant protein will have the same sequence as the wild-type protein and part of the protein will have the sequence of the mutant protein. Furthermore, depending on where the mutation occurs the mutant protein will terminate when the nucleotide sequence codes for a stop codon (indicated as * in the Figures). Thus, different mutant proteins will be produced depending on where the mutation occurs.

Alzheimer's Disease is a representative disease diagnosable and treatable according to the invention. AD is a neurodegenerative disease characterised by idiopathic progressive dementia and is the fourth highest major cause of death in developed countries. It affects 5 to 11% of the population over the age of 65 and as much as 47% of the population over the age of 85. At present there are an estimated 4 million patients suffering from AD in the U.S.A.(see Coleman, Neurobiol. of Ageing, 15, Suppl. 2, pp 577-578, 1994), and an estimated 20 million Alzhemier's patients worldwide.

The clinical criteria for AD diagnosis have been defined (see Reisberg et al., Am. J. Psych. 12, pp 1136-1139, 1982; McKhann *et al.*, Neurology, 34, pp 939-944, 1984). The early symptoms of AD vary but generally include depression, paranoia and anxiety. There is also a slow degeneration of intellectual function and memory. In particular, cognitive dysfunction and specific disturbances of speech (aphasia), motor activity (apraxia), and recognition of perception (agnosia) can occur.

There is not yet general consensus in a test for *ant mortem* diagnosis for AD due to the lack of knowledge of the pathogenic mechanisms involved in AD. Diagnosis of AD is made by examination of brain tissue. Such diagnosis is usually carried out on individuals *post mortem*. The diagnosis is based on the presence of a large number of intraneuronal neurofibrillary tangles and of neuritic plaques in the brain tissue, in particular in the neocortex and hippocampus. In order to identify the various types of plaques (e.g. neuritic plaques), neuropil threads and neurofibrillary tangles, staining and microscopic examination of several brain tissue sections is necessary. Neuritic plaques are believed to be composed of degenerating axons (e.g., neuropil threads), nerve terminals and possibly astrocytic and microglial elements. It is also often found

that neuritic plaques have an amyloid protein core. The neurofibrillary tangles comprise normal and paired helical filaments and are believed to consist of several proteins.

There are two major types of AD, late onset (>60 years) and early onset (<60 years). Approximately 75% of all AD cases are late onset and only 25% are early onset. Of the latter group 2.4% consists of the familial type of AD linked to chromosome 21 and 28% of the cases are considered to be linked to chromosome 14, as discussed below. In addition, a recent linkage to chromosome 1 has been established for juvenile onset (0.4%). Sporadic cases are the most prominent group (60%).

In the most common late onset group, 10 to 30% of cases are considered to be related to chromosome 19, and in particular the apolipoprotein-E_4 gene is considered to be an important risk factor. The remaining late onset cases are non-familial or "sporadic" cases (see Van Broeckhoven *et al.*, Europ. Neurol., 35, pp 8-19, 1995 and Table 1). For these cases relatively little is known and previously no data was available which suggested a possible cause of AD.

At present, it is unclear whether the formation of neuritic plaques and/or neurofibrillary tangles is directly responsible for causing AD. The formation of neuritic plaques, neuropil threads and/or neurofibrillary tangles may be a consequence of a more fundamental cellular or biochemical change.

Diagnostic methods of the invention will include the detection of nucleic acid sequences, preferably via procedures which involve formation of a nucleic acid duplex between two nucleic acid strands, i.e., a nucleic acid probe and a complementary sequence in DNA from a biological sample, or detection of a protein, preferably a mutant or hybrid wild-type/nonsense protein, as defined herein.

1. Preparation and Detecti n of DNA for Genetic Screening.

Typically, DNA is prepared from the biological sample by DNA extraction procedures well-known in the art (see, e.g., Sambrook et al., 1990, A Laboratory Manual for Cloning, Cold Spring Harbor Press, CSH, NY), and may be further purified if desired, e.g., by electro-elution, prior to analysis.

Methods of detecting a mutated gene in DNA from a biological sample include, but are not limited to the following: (1) polymerase chain reaction (PCR) followed by sizing gel electrophoresis or hybridization with an allele-specific (or sequence-specific) probe; (2) hybridization of the eluted DNA with a nucleic acid probe that is complementary to the mutated gene; (3) allele-specific oligonucleotide (ASO) PCR followed by an amplification-detection system (e.g., gel electrophoresis and staining or HPLC); (4) the ARMS test, in which one primer has a complementary sequence encompassing the mutation which gives rise to the frameshift mutation, and amplification only occurs if the mutated sequence is present.; and (5) nucleotide sequencing.

A DNA probe useful according to the invention is preferably sufficiently complementary to the mutant sequence of the gene so that under stringent conditions the probe only remains bound to the mutant sequence (see Evans *et al.*, Proc. Natl. Acad. Sci. USA, 91:6059-6063, (1994). The probe is preferably labelled using any of the standard techniques known to those skilled in the art, such as radioactively using ^{32}P or any other standard isotopes, or using non-radioactive methods including biotin or DIG labelling. The labelled probe can then be easily detected by methods well known to those skilled in the art.

An alternative method for detecting the presence of the mutant gene is via the polymerase chain reaction (PCR).

Primers having a sequence complementary to the sequence either side of the mutation which gives rise to the frameshift mutation are used to amplify the DNA or RNA (if RNA is being detected a reverse transcriptase stage must be performed, as would be apparent to one skilled on the art) containing the mutation. The mutation in the amplified fragment can then be detected using the probe described above using standard techniques or by sequencing the amplified fragment. The advantages of using the PCR reaction is that the actual mutated sequence is obtained, less starting material is required and the PCR methods allow quantitative as well as qualitative determinations to be made. Quantitative determinations allow the number of copies of a mutated gene present in a particular sample to be estimated, and given this information the severity of the diseased state can be estimated.

Another alternative method for detecting the presence of the mutant gene is one in which one primer has a complementary sequence encompassing the mutation which gives rise to the frameshift mutation. Amplification will therefore only occur if the mutated sequence is present. Newton et al., Nucl. Acids. Res. 17:2503, 1989. The method has previously been used in detecting mutations in the gene responsible for cystic fibrosis, and one skilled in the art could easily perform this test for the detection of the mutant gene of the present invention.

An example of analysis method (1) follows. The DNA is amplified, e.g., using PCR, prior to analysis. Specific conditions for any one PCR, i.e. a PCR targeting a particular sequence, or for any one multiplex PCR, i.e. a PCR targeting a particular set of sequences, may vary but will be known to a person of ordinary skill in the art.

Amplification of a mutated or wild-type nucleic acid sequence can be accomplished directly from an aliquot of the prepared DNA as follows.

25 µl of DNA is aliquotted into a reaction tube containing 25 µl H_2O, 50 µl master mix (see below), 0.5 µl Amplitaq (Perkin Elmer Cetus, Norwalk, CT) and 0.5 µl UNG (Perkin Elmer Cetus, Norwalk, CT). A 50 µl master mix comprises 20 mM Tris HCl, pH 8.3, 100 mM KCl, 5 mM $MgCl_2$, 0.02 µmoles each of dATP, dGTP, dCTP, 0.04 µmoles of dUTP, 20 pmoles of each primer (Perkin Elmer Cetus, Norwalk, CT), and 25 µg gelatin.

A fragment characteristic of the selected amplification sequence can then be visualized under ultraviolet light after ethidium bromide staining a 13% polyacrylamide gel in which an aliquot of the amplification has been electrophoresed. Alternatively, hybridization with allele-specific probes can identify the presence of amplified product from either the normal and/or mutant alleles.

2. Preparation and Detection of Protein for Genetic Screening.

Where the biological molecule to be analyzed is a protein, it may be desirable to release the nucleic acid from biological sample cells prior to protein elution, or to remove nucleic acid from the sample eluate prior to protein analysis. Thus, the sample or eluate may first be treated to release or remove the nucleic acid by mechanical disruption (such as freeze/thaw, abrasion, sonication), physical/chemical disruption, such as treatment with detergents (e.g., Triton, Tween, or sodium dodecylsulfate), osmotic shock, heat, enzymatic lysis (lysozyme, proteinase K, pepsin, etc.), or nuclease treatment, all according to conventional methods well known in the art.

Where a biological sample includes a mutant protein, the presence or absence of which is indicative of a genetic disease, the protein may be detected using conventional detection assays, e.g., using protein-specific probes such as an antibody probe. Similarly, where a genetic disease correlates with the presence or absence of an amino acid or sequence of amino acids, these amino acids may be detected using conventional means, e.g., an antibody which is specific for the native or mutant sequence (see Table 7 for examples of amino acid sequences present in mutant proteins).

Any of the antibody reagents useful in the method of the present invention may comprise whole antibodies, antibody fragments, polyfunctional antibody aggregates, or in general any substance comprising one or more specific binding sites from an antibody. The antibody fragments may be fragments such as Fv, Fab and $F(ab')_2$ fragments or any derivatives thereof, such as a single chain Fv fragments. The antibodies or antibody fragments may be non-recombinant, recombinant or humanized. The antibody may be of any immunoglobulin isotype, e.g., IgG, IgM, and so forth.

In addition, aggregates, polymers, derivatives and conjugates of immunoglobulins or their fragments can be used where appropriate.

The immunoglobulin source for an antibody reagent can be obtained in any manner such as by preparation of a conventional polyclonal antiserum or by preparation of a monoclonal or a chimeric antibody. Antiserum can be obtained by well-established techniques involving immunization of an animal, such as a mouse, rabbit, guinea pig or goat, with an appropriate immunogen.

Preparation of Antibodies

1. Polyclonal antibodies.

The peptide or polypeptide may be conjugated to a conventional carrier in order to increases its immunogenicity, and antisera to the peptide-carrier conjugate is raised in rabbits. Coupling of a peptide to a carrier protein and immunizations are performed as described (Dymecki, S.M., et al., J. Biol. Chem 267:4815-4823, 1992). Rabbit antibodies against this peptide are raised and the sera titered against peptide antigen by ELISA or alternatively by dot or spot blotting (Boersma and Van Leeuwen, 1994, Jour. Neurosci. Methods 51:317. At the same time, the antisera may be used in tissue sections. The sera is shown to react strongly with the appropriate peptides by ELISA, following the procedures of Green et al., Cell, 28, 477-487 (1982). The sera exhibiting the highest titer is used in subsequent experiments.

2. Monoclonal antibodies.

Techniques for preparing monoclonal antibodies are well known, and monoclonal antibodies of this invention may be prepared using a synthetic peptide, preferably bound to a carrier, as described by Arnheiter et al., Nature, 294, 278-280 (1981).

Monoclonal antibodies are typically obtained from hybridoma tissue cultures or from ascites fluid obtained from animals into which the hybridoma tissue was introduced. Nevertheless, monoclonal antibodies may be described as being "raised to" or "induced by" the synthetic peptides or their conjugates.

Particularly preferred immunological tests rely on the use of either monoclonal or polyclonal antibodies and include enzyme linked immunoassays (ELISA), immunoblotting, immunoprecipitation and radioimmunoassays. See Voller, A., Diagnostic Horizons 2:1-7, 1978, Microbiological Associates Quarterly Publication, Walkersville, MD; Voller, A. et al., J. Clin. Pathol. 31:507-520 (1978); U.S. Reissue Pat. No. 31,006; UK Patent 2,019,408; Butler, J.E., Meth. Enzymol. 73:482-523 (1981); Maggio, E. (ed.), Enzyme Immunoassay, CRC Press, Boca Raton, FL, 1980) or radioimmunoassays (RIA) (Weintraub, B., Principles of radioimmunoassays, Seventh Training Course on Radioligand Assay Techniques, The Endocrine Society, March 1986, pp. 1-5, 46-49 and 68-78). For analyzing tissues for the presence of the mutant protein of the present invention, immunohistochemistry techniques are preferably used. It will be apparent to one skilled in the art that the antibody molecule will have to labelled to facilitate easy detection of mutant protein. Techniques for labelling antibody molecules are well known to those skilled in the art (see Harlour and Lane, Antibodies, Cold Spring Harbour Laboratory, pp 1-726, 1989).

Alternatively, sandwich hybridization techniques may be used, e.g., an antibody specific for a given protein. In addition, an antibody specific for a haptenic group conjugated to the binding protein can be used. Another sandwich detection system useful for detection is the avidin or streptavidin system, where a protein specific for the detectable protein has been modified by addition of biotin. In yet another embodiment, the antibody may be replaced with

a non-immunoglobulin protein which has the property of binding to an immunoglobulin molecule, for example Staphylococcal protein A or Streptococcal protein G, which are well-known in the art. The protein may either itself be detectable labeled or may be detected indirectly by a detectable labeled secondary binding protein, for example, a second antibody specific for the first antibody. Thus, if a rabbit-anti-hybrid wild-type/nonsense protein antibody serves as the first binding protein, a labeled goat-anti-rabbit immunoglobulin antibody would be a second binding protein.

In another embodiment, the signal generated by the presence of the hybrid wild-type/nonsense protein is amplified by reaction with a specific antibody for that fusion protein (e.g., an anti-ß-galactosidase antibody) which is detectably labeled. One of ordinary skill in the art can devise without undue experimentation a number of such possible first and second binding protein systems using conventional methods well-known in the art.

Alternatively, other techniques can be used to detect the mutant proteins, including chromatographic methods such as SDS PAGE, isoelectric focusing, Western blotting, HPLC and capillary electrophoresis.

Identification of Diseases According to the Invention

The invention provides methods for identifying diseases caused by or associated with at least one gene having one or more somatic mutations which give rise to a frameshift mutation.

Diseases are identified according to the invention as follows. The nucleotide sequence of a gene suspected of being involved in the pathogenesis of a disease is provided, e.g., from published gene sequences or from cloning and sequencing of a suspect gene. The amino acid sequence

encoded by the gene is then predicted, as are amino acid sequences of encoded mutant proteins. Mutant protein sequences are predicted in +1 and +2 reading frames following a hypothesized frameshift mutation. The location of the frameshift mutation may be hypothesized with respect to certain nucleotide sequence motifs, examples of such motifs including but not limited to GAGA, for example, GAGAG, GAGAC, GAGAA, and GAGAT.

A probe is then prepared that is specific for the mutant protein or an immunogenic fragment thereof (such probes being described hereinabove for detection of proteins or protein fragments). Depending on where the mutation that leads to the frameshift occurs, part of the mutant protein will have the same sequence as the wild-type protein and part of the protein will have the sequence of the mutant protein. Furthermore, depending on where the mutation occurs the mutant protein will terminate when the nucleotide sequence codes for a stop codon (indicated as * in the Figures). Thus, different mutant proteins will be produced depending on where the mutation occurs.

The simplest method of probing for the presence of a particular mutant protein is to make an antibody to that protein or an immunogenic portion thereof. An immunogenic fragment may be synthesized corresponding to the C-terminus of the predicted mutant proteins because even if the mutation occurred at another position in the sequence the probability that the derived mutant protein contains the peptide sequence is increased. Furthermore, the C-terminal region of a protein is more likely to form an epitope than other regions of the protein.

Once a probe is made, a biological sample from a patient having the disease and a biological sample from a patient not having the disease is probed for the presence or absence of the mutant protein, also as described above. Alternatively, several probes may be prepared and the

combination of probes used to probe the tissue sample. The presence of the mutant protein in a biological sample from a patient having the disease and the absence of said mutant protein in a biological sample from a patient not having the disease indicates that the mutant protein is a marker for the disease or susceptibility to the disease.

Treatment of Diseases According to the Invention

The invention also relates to methods for preventing and/or treating diseases, vectors for preventing and/or treating the diseases, and compositions such as nucleic acid sequences and proteins for preventing and/or treating the diseases, which methods and compositions are useful in gene and protein therapies.

The invention includes methods of treatment and/or prevention of a disease caused by or associated with a gene having a somatic mutation giving rise to a frameshift mutation in which a vector comprising an expressible gene encoding a repair enzyme or a ribozyme is administered to a patient suffering from or susceptible to the disease. Preferred diseases which are treated according to the invention include but are not limited to cancer or a neurodegenerative disease, especially AD, the preferred mutant genes of the present invention are those encoding the ß amyloid precursor protein, the Tau protein, ubiquitin, apolipoprotein-E_4 (Apo-E_4), microtubule associated protein II (MAP 2) and the neurofilament proteins, having a deletion, insertion or other modification leading to a frameshift mutation.

Examples of genes encoding such repair enzymes include but are not limited to genes encoding MSH 1-6, PMS 1-2, MLH 1 or GTBP.

The invention includes methods of treatment and/or prevention of a disease in which a vector comprising an

expressible gene encoding the wild-type version of the mutated gene is administered to a patient suffering from or susceptible to the disease. Examples of genes encoding a wild-type version of a mutated gene include but are not limited to those disclosed herein and sequences disclosed in the figures.

A vector comprising an expressible gene encoding a repair enzyme (that is, the wild-type version of a defective repair enzyme which has lead to or contributed to the presence of the mutated gene which causes or is associated with a disease), a ribozyme, or a wild-type version of a mutated gene are characterized as described below.

The vectors of the present invention are preferably nucleic acid vectors comprising RNA or DNA. The vectors may be of linear or circular configuration and may be adapted for episomal or integrated existence in the host cell, as set out in the extensive body of literature known to those skilled in the art. The vectors may be delivered to cells using viral or non-viral delivery systems. The choice of delivery system will depend on whether the nucleic acid sequence to be delivered is to be incorporated into the cell genome or is to remain episomal.

Vectors of the present invention additionally may comprise further control sequences such as enhancers or locus control regions (LCRs), in order to lead to more controlled expression of the encoded gene or genes. LCRs are described in EP-A-0332667. The inclusion of a locus control region (LCR), is particularly preferred as it ensures the DNA is inserted in an open state at the site of integration, thereby allowing expression of the gene or genes contained in the vector. The vectors of the present invention have wide range of applications in *ex vivo* and *in vivo* gene therapy.

The invention also includes treatment of a disease by administration of a pharmaceutical composition comprising the wild-type analog of a mutant protein in admixture with a pharmaceutically acceptable carrier.

A pharmaceutical composition according to the invention will include a therapeutically effective amount f the wild-type analog of the mutant protein, a repair enzyme, or a ribozyme, in admixture with a carrier. A therapeutically effective amount is considered that amount which, when administered to a patient, provides a therapeutic benefit to the patient. Such amounts will generally be in the range of 10ug-100mg of therapeutic protein/kg body weight of the patient, preferably 50ug-10mg, and most preferably 100ug-1mg.

Animal Models for Disease Diagnosis and Treatment According to the Invention

The invention also includes cell lines and transgenic animals for use as disease models for testing or treatment.

A cell line or transgenic animal according to the invention will contain a recombinant gene, also known herein as a transgene, having one or more somatic mutations giving rise to a frameshift mutation which causes or is associated with a disease.

The recombinant gene will encode a mutated protein found to be indicative of a disease. Preferably, the mutant protein will contain an antigenic epitope specific for the diseased state. The recombinant gene may encode a protein comprising at least part of the sequence designated +1 or +2 in any one of Figures 2 to 9, or an immunologically equivalent fragment thereof.

A cell line containing a transgene encoding a mutant protein, as described herein, is made by introducing the transgene into a selected cell line according to any one of several procedures known in the art for introducing a foreign gene into a cell.

A transgenic animal containing such a transgene includes a rodent, such as a rat or mouse, or other mammals, such as a goat, a cow, etc. and may be made according to procedures well-known in the art.

Transgenic animals are useful according to the invention as disease models for the purposes of research into diseases caused by or associated with at least one gene having one or more somatic mutations giving rise to a frameshift mutation, and therapies therefore. By specifically expressing one or more mutant genes, as defined above, the effect of such mutations on the development of the disease can be studied. Furthermore, therapies including gene therapy and various drugs can be tested on the transgenic animals.

Transgenic animals of the present invention may be additionally defective in nucleic acid repair mechanisms. Thus, the mutated gene specifically expressed in the transgenic animal will not be repaired.

Example 1

Described below is an embodiment of the invention involving identification of somatic frameshift mutations in genes encoding proteins which are present in neuronal tissue, and how such mutations are useful in diagnosis of certain disease states.

The DNA sequences coding for the human amyloid A4 protein, Tau, ubiquitin, apolipoprotein E4, MAP 2 and the neurofilament subunits low, medium and high, were obtained from various gene sequence databases.

Using the sequence data, the various GAGA motifs in the sequences were identified, and deletions were hypothesized and the sequences of the derived mutant proteins predicted, as shown in Figures 2 to 9. Both the sequences of the +1 and +2 frameshift mutant proteins were predicted.

By examining the sequences of the hypothesized mutant proteins, a peptide corresponding to the C-terminus of the hypothesized mutant proteins was synthesized. The peptides were synthesized using standard techniques known to those skilled in the art. The peptides having the following sequences were synthesized: RGRTSSKELA; HGRLAPARHAS; YADLREDPDRQ; RQDHHPGSGAQ; GAPRLPPAQAA; KTRFQRKGPS; PGNRSMGHE; EAEGGSRS; and VGAARDSRAA.

Depending on where the mutation that leads to the frameshift occurs, part of the mutant protein will have the same sequence as the wild-type protein and part of the protein will have the sequence of the mutant protein. Furthermore, depending on where the mutation occurs the mutant protein will terminate when the nucleotide sequence codes for a stop codon (indicated as * in the Figures). Thus different mutant proteins will be produced depending on where the mutation occurs.

It is predicted that mutations will occur at GAGA motifs and the sequences of the mutant proteins predicted accordingly.

Peptides were synthesized corresponding to the C-terminus of the predicted mutant proteins because even if the mutation occurred at another position in the sequence the probability that the derived mutant protein contains the peptide sequence is increased. Furthermore, the C-terminus region of a protein is more likely to form an epitope than other regions of the protein.

The uniqueness of the synthesized peptides was confirmed by a gene sequence database search.

Each synthesized peptide was then injected into a rabbit and an antibody having affinity for the peptide purified. The techniques used to obtain the antibodies are standard techniques known to those skilled in the art.

The antibodies obtained were then tested on autopsy material of frontal cortex, temporal lobe and hippocampus of neuropathologically confirmed AD cases and control non-AD cases. The presence of the antibodies is determined using standard detection methods known to those skilled in the art.

Figure 1 shows the presence of the ß amyloid precursor mutant protein in the frontal cortex of an Alzheimer patient identified using an antibody against a peptide predicted by the +1 reading frame of ß amyloid precursor protein. The antibody used had affinity for a peptide having the following sequence RGRTSSKELA.

The results of other immunoreactive tests performed using the antibodies against the predicted peptides are shown in Table 2 and 3. Tables 2 and 3 also indicates the

neuropathological state (presence of plaques/tangles) of the various tissue samples.

It can be seen that the presence of the mutant protein can be detected and correlates with the subject having AD. The presence of one or more of the mutant proteins can therefore be seen to be indicative of AD.

Table 4 summaries the immunoreactivity results within the frontal cortex (area 11), temporal cortex (area 38) and the hippocampus.

Other diseases also may be correlated with the presence of mutant proteins, as defined herein. For example, six patients with Down's syndrome were tested according to the invention. Down's syndrome is trisomy of chromosome 21 which leads to over-expression of ß-amyloid precursor protein. We noted that the frontal and temporal cerebral cortex and hippocampus of these patients contained plaques and neurofibrillary tangles, and hypothesized that such over-expression may promote accumulation of somatic mutations in neurons, by frameshift mutations at a GAGAG motif in the over-expressed ß-amyloid gene. After immunocytochemical staining of tissue from frontal and temporal cerebral cortex from the Down's patients with the above-described antibody specific for the amyloid +1 carboxy terminal peptide, immunoreactivity was observed in the neurofibrillary tangles in 4 of 6 patients. Staining was absent in the frontal cortex of the matched controls. Therefore, the mutant amyloid protein is correlated not only with Alzheimer's disease, but also with other diseases, such as Down's, involving Alzheimer's neuropathology.

It has been found that a number of the mutations occur at GAGA motifs. Table 5 shows the presence of GAGA motifs in various genes of the neuronal system. The motif or, as can be seen from the sequences of Tau and apolipoprotein E4, similar motifs such as GAGAG GAGAC, GAGAA, and GAGAT may be

associated with the frameshift mutations that lead to or are associated with the disease. The presence of the motif or similar motifs in other genes may indicate that they are relevant to a disease. It is also possible that other mutations occur that are not associated with such motifs but still lead to frameshift mutations that cause or are associated with a disease.

Table 6 shows the presence of GAGAG motifs in particular genes of the neuronal system, namely βAPP, Tau and Ubiquitin. This table also indicates, *inter alia*, the chromosomal location of the genes and the molecular weight of the longest polypeptide forms encoded by the genes and the predicted size of the aberrant +1 peptide with its C-terminus against which the antibodies were raised. These peptides were revealed in a Western blot and also identified with a different antibody recognizing an epitope on the unaffected wild-type N-terminus.

Example 2

Selection of Antigenic Peptide

Synthetic polypeptides corresponding in sequence to a portion of a mutant protein (whether such peptides are chemically synthesized or are chemically or recombinantly-generated fragments of a protein), as described herein, will be useful according to the invention as antigenic peptides for generation of antibodies specific for a mutant protein, provided they possess the following characteristics. The synthetic peptide will include a minimum of 12 and preferably 15 amino acid residues, and an optimum length of 20-21 amino acids. The hydrophilicity and antigenic index of the amino acid sequence of the hybrid wild-type/nonsense protein may be determined by Analytical Biotechnology Sciences, Boston, MA, using computer programming. Potential synthetic peptides useful according to the invention include

a stretch of 12-20 amino acids preferably within the carboxy terminal 100-150 amino acids of the hybrid wild-type/nonsense protein.

The amino acid sequence of a selected peptide is searched in a computer database of sequences (e.g., GenBank) to preclude the possibility that at reasonable concentrations, antisera to any one peptide would specifically interact with any protein of a known sequence. Preferred sequences are those which are determined not to have a close homolog (i.e., "close" meaning 80-100% identity).

Example 3

Detection of "Mutant" Protein

Another embodiment of this invention relates to an assay for the presence of the "mutant" or mutant protein in a given tissue as indicative of a disease state. Here, an above-described antibody is prepared. The antibody or idiotype-containing polyamide portion thereof is then admixed with candidate tissue and an indicating group. The presence of the naturally occurring amino acid sequence is ascertained by the formation of an immune reaction, as signaled by the indicating group. Candidate tissues include any tissue or cell line or bodily fluid to be tested for the presence of the mutant protein, as described hereinabove.

Expression of a given hybrid wild-type/nonsense protein may be investigated using antiserum prepared in rabbits against a peptide corresponding to a carboxy terminal stretch of amino acids in the hybrid wild-type/nonsense protein as follows.

CMK cells or U3T3 cells are metabolically labeled with

^{35}S-methionine and extracts are immunoprecipitated with antiserum. If the hybrid wild-type/nonsense protein is present in the cells, then a protein species of corresponding molecular weight will be detected in CMK and U3T3 cells. The protein may be localized to the membrane, nucleus or cytoplasm by Western blot analysis of the nuclear, membrane and cytoplasmic fractions, as generally described in Towbin et al., Proc. Natl. Acad. Sci. USA, 76, 4350-4354 (1979). This localization may be confirmed by immunofluorescence analysis to be associated mainly with the plasma membrane.

Metabolic labeling immunoprecipitation, and immunolocalization assays are performed as described previously (Furth, M.E., et al., Oncogene 1:47-58, 1987; Laemmli, U.K., Nature 227:680-685, 1970; Yarden, Y., et al., EMBO J. 6:3341-3351, 1987; Konopka, J.B., et al., Mol. Cell. Biol. 5:3116-3123, 1985). For immunoblot analysis, total lysates are prepared (using Fruth's lysis buffer) (Fruth, M.E., et al., Oncogene, 1:47-58, 1987). Relative protein concentrations are determined with a colorimetric assay kit (Bio-Rad) with bovine serum albumin as the standard. A protein of lysate containing approximately 0.05 mg of protein was mixed with an equal volume of 2 x SDS sample buffer containing 2 mercaptoethanol, boiled for 5 min., fractioned on 10% polyacrylamide-SDS gels (Konopka, J.B., et al., J.Virol., 51:223-232, 1984) and transferred to immunobilon polyvinyldine difluoride (Millipore Corp., Bedford, MA) filters. Protein blots were treated with specific antipeptide antibodies (see below). Primary binding of the specific antibodies may be detected using anti-IgG second antibodies conjugated to horseradish peroxidase and subsequent chemiluminescence development ECL Western blotting system (Amersham International).

For metabolic labeling, 10^6 cells are labeled with 100 μCi of ^{35}S-methionine in 1 ml of Dulbecco's modified Eagles medium minus methionine (Amersham Corp.) for 16h. Immunoprecipitation of protein from labeled cells with

antipeptide antiserum is performed as described (Dymecki, S.M., et al., J. Biol. Chem 267:4815-4823, 1992). Portions of lysates containing 10^7 cpm of acid-insoluble ^{35}S-methionine were incubated with 1 μg of the antiserum in 0.5 ml of reaction mixture. Immunoprecipitation samples were analyzed by SDS-polylarcylamide gel electrophoresis and autoradiography.

For immunolocalization studies, 10^7 CMK cells are resuspended in 1 ml of sonication buffer (60mM Tris-HCl, pH 7.5, 6 mM EDTA, 15 mM EGTA, 0.75M sucrose, 0.03% leupeptin 12mM phenylmethylsulfonyl fluoride, 30 mM 2-mercaptoethanol). Cells are sonicated 6 times for 10 seconds each and centrifuged at 25,000 xg for 10 min at 4°C. The pellet is dissolved in 1 ml of sonication buffer and centrifuged at 25,000 x g for 10 min at 4°C.

The pellet (nucleus fraction) is resuspended in 1 ml of sonication buffer and added to an equal volume of 2 x SDS sample buffer. The supernatant obtained above (after the first sonication) is again centrifuged at 100,000 x g for 40 min at 4°C. The supernatant (cytosolic fraction) is removed and added to an equal volume of 2 x concentrated SDS sample buffer. The remaining pellet (membrane fraction) is washed and dissolved in sonication buffer and SDS sample buffer as described above. Protein samples are analyzed by electrophoresis on 10% polyacrylamide gels, according to the Laemmli method (Konopka, J.B., et al., Mol. Cell. Biol. 5:3116-3123, 1985). The proteins are transferred from the gels on a 0.45-μm polyvinylidine difluoride membrane for subsequent immunoblot analysis. Primary binding of antibodies is detected using anti-IgG second antibodies conjugated to horseradish peroxidase.

For immunohistochemical localization of a given protein, if desired, CMK cells or U3T3 are grown on cover slips to approximately 50% confluence and are washed with PBS (pH 7.4) after removing the medium. The cells are

prefixed for 1 min at 37°C in 1% paraformaldehyde containing 0.075% Triton X-100, rinsed with PBS and then fixed for 10 min with 4% paraformaldehyde. After the fixation step, cells are rinsed in PBS, quenched in PBS with 0.1 and finally rinsed again in PBS. For antibody staining, the cells are first blocked with a blocking solution (3% bovine serum albumin in PBS) and incubated for 1 h at 37°C. The cells are then incubated for 1 h at 37°C with antiserum (1:100 dilution or with preimmune rabbit serum (1:100) (see below). After the incubation with the primary antibody, the cells are washed in PBS containing 3% bovine and serum albumin and 0.1% Tween 20 and incubated for 1 h at 37 C in fluorescein-conjugated donkey anti-rabbit IgGs (Jackson Immunoresearch, Maine) diluted 1:100 in blocking solution.

The coverslips are washed in PBS (pH 8.0), and glycerol is added to each coverslip before mounting on glass slides and sealing with clear nail polish. All glass slides were examined with a Zeiss Axiophot microscope.

Example 4

Biological Sample Analysis

The above methods for detection of a given mutant protein or nucleic acid are applicable to analyses involving tissues, cell lines and bodily fluids suspected of containing the marker protein.

For example, a sample of CNS tissue suspected of being in a diseased state may be analyzed, it having been previously observed according to the invention that tissue of that particular diseased state contains detectable levels of hybrid wild-type/nonsense proteins relative to healthy tissue.

An aliquot of the suspect sample and a healthy control sample are provided and admixed with an effective amount of an antibody specific for the hybrid wild-type/nonsense protein, as herein described, and an indicating group. The admixture is typically incubated, as is known, for a time sufficient to permit an immune reaction to occur. The incubated admixture is then assayed for the presence of an immune reaction as indicated by the indicating group. The relative levels of the hybrid wild-type/nonsense protein in the suspect sample and the control sample are then compared, allowing for diagnosis of a diseased or healthy state in the suspect sample.

The above types of analyzing for the presence of the hybrid wild-type/nonsense protein may, of course, be performed using analysis for the coding mRNA, e.g., via Northern blot or RNA dot blot analysis, both of which are conventional and known in the art.

Disease Treatment According to the Invention

Once a gene containing a frameshift mutation (i.e., a frameshifted gene), or a mutant protein is correlated with a disease state, the disease is treatable according to the invention as follows: by administering to a patient in need thereof a wild-type version of a mutant gene or the corresponding wild-type transcript; by administering the wild-type version of the hybrid wild-type/nonsense protein; by administering enzymes, or sequences encoding such enzymes, having activity that operates to correct nucleotide insertions or deletions (DNA repair enzymes); by administering enzymes which serve to correct frameshifted RNA via splicing, e.g., ribozymes; or by administering oligonucleotides or sequences encoding oligonucleotides to a cell as templates for repair of an insertion or deletion mutation. A patient in need thereof will include a patient exhibiting symptoms of the disease, even those patients suspected of developing the disease, i.e., who are monitored according to the invention by measuring the a tissue sample, e.g., the cerebrospinal fluid, for the presence of frameshifted peptides (e.g. peptides having an amino acid sequence in the +1 or +2 reading frame).

Therefore, it is contemplated according to the invention that the wild type version of the mutant (i.e., frameshifted) gene, or its encoded wild type protein, may be administered to the patient in order to treat the disease associated with the presence of the mutant gene or its encoded protein.

According to the invention, a vector encoding a non-mutated version of a frameshifted gene under the control of a promoter can be delivered to affected or susceptible cells leading to the production of the correct protein in the cell. Without being bound by any one theory, it is suggested that increasing the percentage of the correct protein produced in relation to the hybrid wild-

type/nonsense protein will reduce or prevent further progression of the disease, and possibly reverse the diseased state. For example, in neurons having undergone a frameshift mutation in one of two expressed alleles, the balance between the wild-type and mutant transcripts may be shifted in favor of the wild-type one as a result of the treatments described herein. In addition, it is possible that not every mutant gene and transcript results in a mutant protein that is directly toxic to the neuronal tissue. For example, the mutant protein may be routed to the lysosomal system or just secreted (e.g. by the constitutive or regulated pathway)and degraded elsewhere. However, sometimes the mutant protein will be accumulated in the membranes of organelles, for instance in the endoplasmic reticulum, thus disrupting the normal processes of the cellular machinery. This may be especially true if both alleles of a given gene or if different genes are mutated.

The wild-type version of the mutated gene encodes the correct protein. The wild-type gene corresponds to the mutated gene found present in the affected cells and preferably the wild-type gene is expressed at high levels leading to the production of more wild-type protein than the mutated protein. When the disease is a neurodegenerative disease, preferred wild-type genes include the ß amyloid protein gene, the Tau gene, the ubiquitin gene, the apolipoprotein-E_4 gene, the microtubule associated protein II (MAP2) gene and the neurofilament protein genes. The sequences of these genes are provided herein in the figures. Other preferred wild-type genes include the alpha and beta tubulin genes, the sequences of which are found in Cowan et al., Mol. and Cellular Biology, 3, 1738-1745(1983) and Lewis et al., J. Mol. Biol. 182, 11-20(1985), repectively.

When the disease is a non-hereditary cancer, preferred wild-type gene sequences include but are not limited to the human p53 gene and the bcl-2 gene. Mammalian phosphoprotein

p53 has been shown to play an essential role in regulation of cell division and is required for the transition from phase G0 to G1 of the cell cycle. P53 is normally present in very low levels in normal cells and is believed to be a tumor suppressor gene; when present at high levels, p53 has been shown to play a role in transformation and malignancy. P53 gene alleles from normal and malignant tissues have been shown to contain BglII site polymorphism (Buchman et al., 1988, Gene 70:245). The p53 coding region contains several GAGA motifs, e.g., GAGAC at position 1476 of the sequence published in Buchman et al., GAGA at position 1498; GAGA at position 1643; and GAGA at position 1713, which motifs present candidate sites for frameshift mutations according to the invention. A frameshift mutation within the p53 gene thus may lead to loss of the natural p53 tumor suppressor function. Detection of such a mutation in p53 may be diagnostic of pre-malignancy or malignancy, and treatment as described herein which results in correction of p53 function may restore tumor suppressor function.

It is also contemplated according to the invention that a mutant gene, which mutation comprises a somatic mutation giving rise to a frameshift mutation that includes a nucleotide deletion or insertion (i.e., of one or more nucleotides), may be due to a deficit in one or more DNA repair enzymes and/or associated proteins in the cell containing the mutant gene. Therefore, where the disease is associated with a deficiency in a DNA repair enzyme, the disease may be treated by administering a DNA repair enzyme, or sequences encoding the repair enzyme, to the patient.

The present invention thus further provides a method of treatment and/or prevention of a disease caused by or associated with at least one gene having one or more somatic mutations giving rise to a frameshift mutation, comprising administering a vector encoding one or more repair enzymes under the control of a promoter to a patient suffering from or likely to suffer from the disease, or delivering the

repair enzyme itself to a target cell containing the frameshifted gene.

The repair enzyme can be any repair enzyme which is capable of or contributes to, the repair of a somatic mutation which leads to a frameshift mutation, such as a dinucleotide deletion. Preferred repair enzymes include Mut H, Mut S, Mut L, and Mut U and the mammalian homologues thereof. These include MSH 1-6, PMS 1-2 and MLH 1, as described in Prolla et al., Science 265:1091, 1994; Strand et al., 1993, Nature 365:274; Kramer et al., 1984, Cell 38:879; Johnson et al., 1995, Science 269:238; Horii et al., 1994, Biochem. Biophys. Res. Comm. 204:1257; Jeyaprakash et al., 1994, Mutation Research 325:21; J.H.J. Hoeijmakers, 1987, Journal of Cell Science Suppl. 6:111-125; and Kunkel et al., supra.

The invention also encompasses methods of combatting diseases caused by at least one gene having one or more somatic mutations giving rise to a frameshift mutation by targeting the mRNA transcript transcribed from the mutated gene. Thus, it is also contemplated according to the invention that a frameshift mutation within a gene may be corrected at the level of the frameshifted RNA via splicing using a ribozyme having specificity for the correct mRNA sequence (see Denman et al., Arch. Of Biochem. Biophysics. 323,71-78,1995), and correcting the mutant mRNA using ribozymes. The disease associated with the frameshifted gene is thus treated by administering an appropriate ribozyme, or sequences encoding the ribozyme, to the patient.

Ribozymes of selected specificities may be made as described by Sullenger & Cech,Nature 371: 619-622, 1994), herein incorporated by reference. The ribozyme encoded by the vector is preferably specific for RNA containing the mutations described above or for the RNA encoding a defective repair enzyme. For example, if a defect in a

repair enzyme has led to the mutated gene, by correcting the transcribed message from the repair enzyme gene the function of the repair enzyme can be restored.

Ribozymes and sequences encoding ribozymes may be prepared as described by Tuschl et al., Curr. Opin. Struc. Biol. 5:296, 1995 and Wahl et al., Curr. Opin. Struc. Biol. 5:282, 1995.

The invention also encompasses methods of treating diseases caused by or associated with at least one gene having one or more somatic mutations giving rise to a frameshift mutation by delivery of oligonucleotides or sequences encoding oligonucleotides to a target cell containing a frameshifted gene. The oligonucleotides will have a wild-type sequence with respect to the corresponding mutant sequence of the frameshifted gene, and thus may serve as templates for correction of the mutant sequence. For example, triple helix-forming oligonucleotides may be used to target mutations to selected genes within mammalian cells. Targeted mutagenesis in vivo depends upon the strength and specificity of the third strand binding. Oligonucleotides with strong target site binding affinity, i.e., with full target site homology (except for the base to be corrected, which base will be mismatched between the oligonucleotide and the nuclear DNA) are preferred. Also preferred are oligonucleotides between 10 - 30 nucleotides in length and containing a GAGA, GAGAC, GAGAG, GAGAA or GAGAT motif. Formation of a site-specific triple strand brings the wild type oligonucleotide sequence into proximity with the base pair to be mutated. DNA repair enzyme will then recognize the mismatch and will repair the mismatch. Targeted mutagenesis via triple helix formation has been shown to work in mammalian cells (see Wang et al., 1995, Mol. Cell Biol. 15:1759).

Disease treatment according to the invention is described below and includes preparation of the administered

substance and administration of the substance to a patient suffering from a disease according to the invention. As used herein "substance" refers to any one of the following: a DNA repair enzyme, a ribozyme, a nucleic acid sequence encoding a repair enzyme or a ribozyme, a nucleic acid sequence comprising a wild type version of a mutant gene, and a wild type protein encoded by the wild type gene, an antibody specific for the frameshifted (nonsense) protein, and oligonucleotides having a wild type sequence to serve as a template for repair of a frameshift mutation.

Disease treatment according to the invention may be accomplished as follows. In Example 5, administration of nucleic acid sequences according to the invention is described.

In Example 6, the purification of DNA repair enzymes is described. In Example 7, administration of proteins, ribozymes, or nucleic acids using liposomes is described. And in Example 8, delivery of these substances across the blood-brain barrier is described.

Example 5

Administration of Nucleic Acids

Delivery of nucleic acid sequences encoding substances which effect or facilitate repair of the frameshifted gene is carried out as follows. Sequences encoding repair enzymes, ribozymes, a wild-type version of a mutant gene, or a wild-type oligonucleotide sequence may be cloned into an appropriate vector for expression in a desired mammalian cell. The vector will include a promoter that is expressed in the target cell type, and also may include an enhancer and locus control region, as selected for expression in a given cell type. Examples of vectors useful according to the invention include but are not limited to any vector which results in successful transfer of the coding sequences

to the target mammalian cell, including both viral or non-viral vectors, e.g., retroviral vectors or adenoviral vectors.

For example, the retroviral gene transfer vector SAX (Kantoff et al., Proc. Nat. Aca. Sci. 83:6563, 1986) may be used to insert a selected coding sequence into a target cell. SAX is a moloney virus based vector with the neoR gene promoted from the retroviral LTR and the human ADA gene promoted from an internal SV40 promoter. Thus, the SAX vector may be engineered by one of skill in the art to contain the coding sequence for a DNA repair enzyme, a ribozyme, or a wild-type version of a mutant gene, or a selected oligonucleotide template sequence, identified as described herein, e.g., by substituting the desired coding region for the hADA coding region in the SAX vector.

It has been previously shown that ribozymes can be targeted to and can correct specific errors via a transplicing mechanism. See Sullenger & Cech, Nature, 371, pp 619-622, 1994. By correcting the mutant mRNA, the correct protein will be translated thus preventing and/or treating the disease. One skilled in the art following the teaching of Sullenger & Cech would easily be able to design a ribozyme to correct the mRNA transcript transcribed from the mutant gene of the present invention.

Expression vectors are known in the art which encode, or may be engineered to encode, a selected ribozyme. Yuyama et al., Nucl. Acids Res. 22:5060, 1994, describe a multifunctional expression vector encoding several ribozymes. This vector may be adapted to encoded a ribozyme of a selected specificity by substituting one or both ribozyme sequences in the vector for a selected ribozyme sequence. Zhou et al., Gene 149:33, 1994, and Yamada et al., Virology 205;121, 1994, describe retroviral transduction of ribozyme sequences into T cells. These retroviral vectors may be adapted to encode a selected

ribozyme sequence. Liu et al., Gene Therapy 1:32, 1994, and Lee et al., Gene Therapy 2:377, 1995, describe expression vectors which are adaptable for use in expression of any nucleic acid sequence contemplated according to the invention.

Once the vector contains the desired coding region, the vector may be introduced ex vivo into a selected population of cells isolated from a patient, and the transfected cells then reintroduced into a patient. The coding sequence for the repair enzyme, ribozyme, wild-type version of a mutant protein, or oligonucleotide, will then be expressed in the patient, and the reintroduced population of cells may expand and thus provide a cell population in which the frameshift mutation is corrected.

Alternatively, the vector may be encapsulated in liposomes, as described hereinbelow, and administered to the patient. The vector-containing liposomes may be prepared so as to target a particular cell type suspected or known to contain a frameshift mutation in a particular gene, and the vector will be introduced into that cell type and result in correction of the frameshift mutation, i.e., by virtue of the presence of the encoded DNA repair enzyme or by substitution of the mutant gene sequence with the corresponding wild-type sequence. Should the frameshift mutation result from a splicing defect, the presence of an appropriate ribozyme is expected to result in correction of the splicing defect, and thus of the frameshift mutation.

Example 6

Preparation of DNA Repair Enzymes

DNA repair enzymes may be prepared and purified for use in the invention as follows. A general review of purification methods for DNA repair enzymes can be found in

DNA REPAIR: A Laboratory Manual of Research Procedures, edited by E. Friedberg and P.C. Hanawalt, published by Marcel Dekker, New York.

With regard to purification of repair enzymes useful according to the invention, the first step of the purification process, molecular sieving, serves to separate the DNA repair enzymes from the vast majority of proteins with larger sizes based on relative rates of migration of the DNA repair enzymes and the contaminating proteins through the molecular sieve matrix.

Molecular sieving can be accomplished by many methods, including gel filtration and electrophoresis. In gel filtration proteins flow around and through pores in beads made from dextran, polyacrylamide, agarose, agarose and acrylamide composites, or other material. The size of the bead pores include or exclude proteins based on size. In electrophoresis, proteins move in an applied electric field through a sizing matrix.

The preferred molecular sieving method for use with the present invention is gel filtration because the enzyme can be easily recovered and because the method is independent of such factors as net protein charge. The pore size of the beads used with this method are selected to maximize separation of DNA repair enzymes from the bulk of other proteins. A general guideline for selecting the gel filtration matrix is that the gel should have an exclusion limit greater than about twice the molecular weight or Stokes' radius of the DNA repair enzyme and less than about 60,000 daltons or 35 Angstroms.

A wide variety of elution buffers may be used to elute the DNA repair enzyme from the gel filtration column. The selected buffer should satisfy the following criteria: 1) the buffer should not denature or inactivate the DNA repair enzyme, 2) the buffer should not permit ionic adsorption of

the DNA repair enzyme to the gel filtration media, and 3) the buffer should be compatible with loading of the eluate onto the nucleic acid affinity column, that is, the elution buffer should be chosen so that complexes will form between the DNA repair enzyme and the immobilized nucleic acids of the affinity column.

The second step of the purification process-nucleic acid binding-separates the DNA repair enzymes from the remaining protein impurities by the ability of DNA repair enzymes to reversibly bind to nucleic acids. Separation by nucleic acid binding can be accomplished by various methods, including nucleic acid affinity chromatography. In this method, nucleic acids are immobilized on an inter matrix, such as agarose, polyacrylamide beads, cellulose or other media. Depending on the DNA repair enzyme which is being purified, the immobilized nucleic acids may be double- or single-stranded DNA, double- or single-stranded RNA, or other types, lengths, structures or combination of nucleic acids, such as tRNA, Z-DNA, supercoiled DNA, ultraviolet-irradiated DNA or DNA modified by other agents. Single-stranded DNA is preferred.

The nucleic acids may be attached to the solid phase matrix by a variety of methods, including covalent attachment of the nucleic acid through primary amines or by absorbing the nucleic acids to a matrix such as cellulose, which releases nucleic acids slowly. The preferred immobilization method is use in a cyanogen-bromide activated Sepharose and to bind the nucleic acids to the activated Sepharose covalently. Alternatively, single-stranded DNA covalently bound to agarose can be purchased commercially from Bethesda Research Labs. Gaithersburg, MD. (Catalog No. 5906SA).

The DNA repair enzymes are applied to the nucleic acids in a solution which should satisfy the following criteria: 1) the solution should permit reversible binding

of the DNA repair enzyme to the nucleic acids, 2) the solution should reduce nonspecific binding of contaminating proteins to the nucleic acids, and 3) the solution should not cause damage to the nucleic acids. In general, a neutral buffered solution with physiological saline and 1 mM EDTA will satisfy these criteria. As discussed above, in accordance with the invention, the nucleic acid affinity column. Accordingly, the elution buffer used with the molecular sieve column should be chosen to satisfy the foregoing criteria.

The bound DNA repair enzymes are eluted from the nucleic acid affinity column with a gradient which removes the enzyme from the nucleic acid at a characteristic condition and concentrates the enzyme by the focusing effect of the gradient. The elution system, however, should not denature the enzyme or introduce contaminants into the final product. A gradient of NaCl up to 1.0M will in general be sufficient to reverse binding of most DNA repair enzymes to nucleic acids. In appropriate cases, the gradient may be one of another salt, increasing or decreasing pH, temperature, voltage or detergent, or, if desired, a competing ligand may be introduced to replace the nucleic acid binding.

Example 7

Liposomal Delivery According to the Invention

Substances may be administered according to the invention using any delivery means known in the art. Described below is liposomal delivery. Liposomes which are used to administer the DNA repair enzymes, ribozymes, a wild-type version of a hybrid wild-type/nonsense protein, or nucleic acid sequences encoding enzymes, ribozymes or a wild-type protein, antibodies specific for a hybrid wild-type/nonsense protein or oligonucleotides can be of various types and can have various compositions. The primary

restrictions are that the liposomes should not be toxic to the living cells and that they should deliver their contents into the interior of the cells being treated.

The use of pH sensitive liposomes to mediate the cytoplasmic delivery of calcein and FITC dextran has been described (see Straubinger et al., Cell 32:1069-1079, 1983; and Straubinger et al., FEBS Letters 179:148-154, 1985. Other discussions of pH sensitive liposomes can be found in chapter 11 of the book CELL FUSION, edited by A.E. Sowers, entitled "Fusion of Phospholipid Vesicles Induced by Divalent Cations and Protons" by Nejat Duzgunes et al., Plenum Press, N.Y., 1987, 241-267. See also Ellens et al., Biochemistry, 23:1532-1538, 1984, and Bentz et al., Biochemistry 26:2105-2116, 1987.

The liposomes may be of various sizes and may have either one or several membrane layers separating the internal and external compartments. The most important elements in liposome structure are that a sufficient amount of enzyme or nucleic acid be sequestered so that only one or a few liposomes are required to enter each cell for delivery of the substance, and that the liposome be resistant to disruption. Liposome structures include small unilamellar vesicles (SUVs, less than 250 angstroms in diameter), large unilamellar vesicles (LUVs, greater than 500 angstroms in diameter), and multilamellar vesicles (MLs). In the example presented below, although SUVs are used to administer DNA repair enzymes, the methods are applicable to administration of ribozymes or sequences encoding repair enzymes, ribozymes, antibodies specific for hybrid wild-type/nonsense proteins, or wild-type genes or their encoded proteins, or oligonucleotides.

SUVs can be isolated from other liposomes and unincorporated enzyme by molecular weight can be isolated from other liposomes and unincorporated enzyme by molecular sieve chromatography, which is precise but time consuming

and dilutes the liposomes, or differential centrifugation, which is rapid but produces a wider range of liposome sizes.

The liposomes may be made from natural and synthetic phospholipids, glycolipids, and other lipids and lipid congeners; cholesterol, cholesterol derivatives and other cholesterol congeners; charged species which impart a net charge to the membrane; reactive species which can react after liposome formation to link additional molecules to the liposome membrane; and other lipid soluble compounds which have chemical or biological activity.

The liposomes useful according to the invention may be prepared, for example, as described in U.S. Patent No. 5,296,231, which describes preparation of liposomes containing a DNA repair enzyme, although it should be borne in mind that liposomes useful according to the invention may contain any one of the substances as herein described. Briefly, by combining a phospholipid component with an aqueous component containing the DNA repair enzyme (or desired substance) under conditions which will result in vesicle formation. The phospholipid concentration must be sufficient to form lamellar structures, and the aqueous component must be compatible with biological stability of the enzyme. Methods for combining the phospholipids onto glass and then vesicles will form include: drying the phospholipids onto glass and then dispersing them in the aqueous component; injecting phospholipids dissolved in a vaporizing or non-vaporizing organic solvent into the aqueous component which has previously been heated; and dissolving phospholipids in the aqueous phase with detergents and then removing the detergent by dialysis. The concentration of the DNA repair enzyme in the aqueous component can be increased by lyophilizing the enzyme onto dried phospholipids and then rehydrating the mixture with a reduced volume of aqueous buffer. SUV's can be produced from the foregoing mixtures either by sonication or by dispersing the mixture through either small bore tubing or

through the small orifice of a French Press.

DNA repair enzymes incorporated into liposomes can be administered to living cells internally or topically. Internal administration to animals or humans requires that the liposomes be pyrogen-free and sterile. To eliminate pyrogens, pyrogen-free raw materials, including all chemicals, enzymes, and water, are used to form the liposomes. Sterilization can be performed by filtration of the liposomes through 0.2 micron filters. For injection, the liposomes are suspended in a sterile, pyrogen-free buffer at a physiologically effective concentration. Topical administration also requires that the liposome preparation be pyrogen-free, and sterility is desirable. In this case, a physiologically effective concentration of liposomes can be suspended in a buffered polymeric glycol gel for even application to the skin. In general, the gel should not include non-ionic detergents which can disrupt liposome membranes. Other vehicles can also be used to topically administer the liposomes.

The concentration of the substance in the final preparation can vary over a wide range, a typical concentration being on the order of 50 ug/ml. In the case of pH sensitive liposomes, lower concentrations of the substance can be used, e.g., on the order of 0.01 to 1.0 ug/ml for liposomes administered to cells internally. In case of topical application, higher liposome concentrations used, e.g., ten or more times higher.

Example 8

Adminstration Across the Blood-Brain Barrier

Where it is desired according to the invention to administer a DNA repair enzyme, a ribozyme, or their nucleic acid coding sequences, a wild-type version of a hybrid wild-type/nonsense protein that is associated with the disease,

or its coding sequence, or oligonucleotides or their coding sequences, or liposomes containing such substances, to an individual such that the administered material crosses the blood-brain barrier, several methods are known in the art.

For example, a substance to be administered, whether it be protein or nucleic acid or liposome, may be co-administered with a polypeptide, for example a lipophilic polypeptide that increases permeability at the blood-brain barrier. Examples of such polypeptides include but are not limited to bradykinin and receptor mediated permeabilizers, such as A-7 or its conformational analogues, as described in U.S. Patent Nos. 5,112,596 and 5,268,164. The permeabilizing polypeptide allows the co-administered repair enzyme, ribozyme, coding sequence or liposome to penetrate the blood-brain barrier and arrive in the cerebrospinal fluid compartment of the brain, where the repair enzyme, ribozyme, or coding sequence may then reach and enter a target neuronal cell. Alternatively, the substance to be administered may be coupled to a steroidal estrogel or androgel to increase binding to steroid receptors and thus access to the brain.

Another exemplary method for administering a substance such as a DNA repair enzyme, ribozyme, antibody, nucleic acids, or liposomes containing such molecules, according to the invention includes forming a complex between the substance to be administered and an antibody that is reactive with a transferrin receptor, as described in U.S. Patent No. 5,182,107. The complex may include a cleavable or non-cleavable linker and is administered under conditions whereby binding of the antibody to a transferrin receptor on a brain capillary endothelial cell occurs and the substance is transferred across the blood-brain barrier in active form.

Other Dosages and Modes of Administration

A patient that is subject to a disease state which is associated with a frameshift mutation may be treated in accordance with the invention, as described above, via in vivo, ex vivo or in vitro methods. For example in in vivo treatments, a nucleic acid vector encoding a repair enzyme, ribozyme, wild-type version of a hybrid wild-type/nonsense protein, or oligonucleotide that corresponds in sequence (except for the inserted or deleted nucleotide(s)) to the frameshifted region of the hybrid wild-type/nonsense protein coding region can be administered to the patient, preferably in a pharmaceutically acceptable delivery vehicle and a biologically compatible solution, by ingestion, injection, inhalation or any number of other methods. The dosages administered will vary from patient to patient; an "effective dose" will be determined by the level of enhancement of function of the transferred genetic material balanced against any risk of deleterious side effects. Monitoring gene expression and/or the presence or levels of the encoded mutant protein or its corresponding "sense" protein will assist in selecting and adjusting the dosages administered. Generally, a composition including a vector will be administered in a single dose in the range of 10 ng - 100 ug/kg body weight, preferably in the range of 100 ng - 10 ug/kg body weight, such that at least one copy of the sequence is delivered to each target cell. A composition including a protein, e.g., a DNA repair enzyme or a wild-type version of a hybrid wild-type/nonsense protein, will be administered in single or multiple doses, as determined by the physician, in the range of 10 ug - 1 mg, or within the range of 100 ug - 50 ug. A composition including a nucleoprotein such as a ribozyme will be administered in single or multiple doses, as determined by the physician, in the range of 50 ug - 1 mg, or within the range of 100 ug - 500 ug. A composition including an oligonucleotide will be administered in a single dose in the range of 5 ng - 10 ug, or within the range of 100 ng - 500 ng. Any of the above

dosages may be administered according to the body weight of the patient, as determined by the physician.

Ex vivo transduction is also contemplated within the present invention. Cell populations can be removed from the patient or otherwise provided, transduced with a vector in accordance with the invention, then reintroduced into the patient. The number of cells reintroduced into the patient will depend upon the efficiency of vector transfer, and will generally be in the range of 10^4 - 10^6 transduced cells/patient.

The cells targeted for ex vivo gene transfer in accordance with the invention include any cells to which the delivery of the vector is desired, for example, neuronal cells or stem cells.

Protein, nucleic acid, or cells administered according to the invention is preferably administered in admixture with a pharmaceutically acceptable carrier substance, e.g., magnesium carbonate, lactose, or a phospholipid to form a micelle, the carrier and protein, nucleic acid or cell together can form a therapeutic composition, e.g., a pill, tablet, capsule or liquid for oral administration to the mammal. Other forms of compositions are also envisioned, e.g., a liquid capable of being administered nasally as drops or spray, or a liquid capable of intravenous, parenteral, subcutaneous, or intraperitoneal administration. The substance administered may be in the form of a biodegradable sustained release formulation for intramuscular administration. For maximum efficacy, where zero order release is desirable, e.g., an implantable or external pump, e.g., an Infusaid™ pump (Infusaid Corp, MA), may be used.

Kits Useful According t the Inventi n

The invention encompasses kits for diagnosis or treatment of diseases according to the invention, a kit including suitable packaging materials and one or more of the following reagents: a nucleic acid probe is as defined hereinabove, and optionally means for detecting the probe when bound to its complementary sequences. For example, the nucleic acid probe may be labeled, e.g., radiolabeled, fluorescently labeled, etc., or may be detected via indirect labeling techniques, e.g., using a biotin/avidin system, well known in the art.

A diagnostic system, preferably in kit form, comprises yet another embodiment of this invention. This system is useful for assaying the presence of a hybrid wild-type/nonsense protein or its derivative in cells by the formation of an immune complex. This system includes at least one package that contains an antibody of this invention. Optionally, a kit also may include a positive tissue sample control.

Antibodies are also utilized along with an "indicating group" also sometimes referred to as a "label". The indicating group or label is utilized in conjunction with the antibody as a means for determining whether an immune reaction has taken place, and in some instances for determining the extent of such a reaction.

The terms "indicating group" or "label" are used herein to include single atoms and molecules that are linked to the antibody or used separately, and whether those atoms or molecules are used alone or in conjunction with additional reagents. Such indicating groups or labels are themselves well-known in immunochemistry and constitute a part of this invention only insofar as they are utilized with otherwise novel antibodies, methods and/or systems.

For example, an antigen-specific antibody or antibody fragment is detectably labeled by linking the same to an enzyme and use it in an EIA, or enzyme-linked immunosorbent assay (ELISA). This enzyme, in turn, when later exposed to a substrate in such a manner as to produce a chemical moiety which can be detected, for example, by spectrophotometric, flourometric or, most preferably, by visual means. The substrate may be a chromogenic substrate which generates a reaction product visible to the naked eye.

Enzymes which can be used to detectably label the binding protein which is specific for the desired detectable mutant protein, include, but are not limited to, alkaline phosphatase, horseradish peroxidase, glucose-6-phosphate dehydrogenase, staphylococcal nuclease, delta-V-steroid isomerase, yeast alcohol dehydrogenase, alpha-glycerophosphate dehydrogenase, triose phosphate isomerase, asparaginase, ribonuclease, urease, catalase, glucose-6-phosphate dehydrogenase, glucoamylase and acetylcholinesterase.

By radioactively labeling the binding protein, for example, the antibody, it is possible to detect the antigen bound to a solid support through the use of a radioimmunoassay (RIA). The radioactive isotope can be detected by such means as the use of a gamma counter or a scintillation counter or by autoradiography. Isotopes which are particularly useful for the purpose of the present invention are: ^{3}H, ^{131}I, ^{14}C, and preferably ^{125}I.

It is also possible to label the first or second binding protein with a fluorescent compound. When the fluorescently labeled antibody is exposed to light of the proper wave length, its presence can then be detected due to fluorescence. Among the most commonly used fluorescent labelling compounds are fluorescein isothiocyanate, rhodamine, phycoerythrin, phycocyanin, allophycocyanin, o-phthaldehyde and fluorescamine.

The first or second binding protein also can be detectably labeled by coupling it to a chemiluminescent compound. The presence of the chemiluminescent-tagged antibody is then determined by detecting the presence of luminescence that arises during the course of a chemical reaction. Examples of particularly useful chemiluminescent labeling compounds are luminol, isoluminol, theromatic acridinium ester, imidazole, acridinium salt and oxalate ester.

Likewise, a bioluminescent compound may be used to label the first or second binding protein. Bioluminescence is a type of chemiluminescence found in biological systems in which a catalytic protein increases the efficiency of the chemiluminescent reaction. The presence of a bioluminescent protein is determined by detecting the presence of luminescence. Important bioluminescent compounds for purposes of labeling are luciferin, luciferase and aequorin.

The invention also includes diagnostic reagents f r use in the present invention, such as nucleic acid sequences, probes and antibody molecules, and/or positive tissue controls, as described above, and kits including such reagents for use in diagnosing or treating a disease.

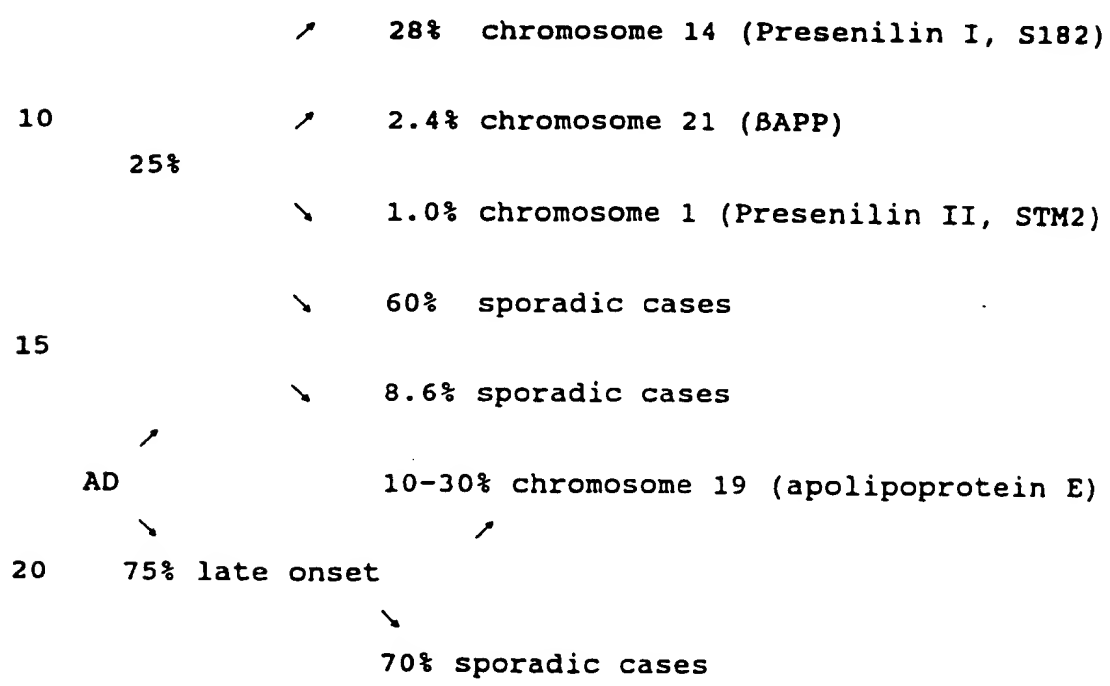
An indicating group or label is preferably supplied along with the antibody and may be packaged therewith or packaged separately. Additional reagents such as hydrogen peroxide and diaminobenzideine, and nickel ammonium sulfate may also be included in the system when an indicating group such as HRP is utilized. Such materials are readily available in commerce, as are many indicating groups, and need not be supplied along with the diagnostic system. In addition, some reagents such as hydrogen peroxide decompose on standing, or are otherwise short-lived like some radioactive elements, and are better supplied by the end-user.

Other Embodiments

It will be understood that the invention is described by way of illustration only. Many other embodiments of the present invention in addition to those herein described will be apparent to those skilled in the art from the descripti n herein given without departing from the scope of the present invention as defined in the appended claims.

Table 1

Table 1. Estimated percentages of early (<60 years) and late (>60 years) Alzheimer's disease (AD) onset.
60% non-familial/40% familial (Van Broeckhoven, 1995, Eur. Neurol. 35, 8-19)



According to Ott et al. (Br. Med. J. 310, 970-973, 1995) the prevalence of dementia increases exponentially with age. In a case finding study in a general population of subjects of 65 years of age and older, 34% of the subjects aged 85 and older had dementia. Of all cases of dementia, 72% were cases of Alzheimer's disease.

Table 2

Table 2. Immunoreactivities within the frontal lobe of the human frontal cortex for various neuronal proteins of which the mRNA is expressed in the +1 reading frame. Tissues were obtained from controls and neuropathologically confirmed Alzheimer cases.

autopsy no.	age (years)	sex (m/f)	H^{+1}	$MAP2^{+1}$	AMY^{+1}	Tau^{+1}	Ubi^{+1}	N	F	-
controls										
89004	34	m	-		-	-		-	-	
81267	43	m	-		-	-		-	-	
88149	58	m	-		-	-		-	-	
90183	65	f	-		-	-		-	-	
90202	72	m	-		-	-		-	-	
91090	80	f	-		-	-		+	+	
91126	82	f	-		-	-		-	-	
90203	85	m	-		-	-		-	-	
88109	85	f	-		-	-		-	-	
81033	90	f	-		-	-		-	+	
90206	90	f	-		-	-		-	-	
% pos. staining			0%		0%	0%		9%	[illegible]%	
Alzheimer's cases										
89166	40	m	-		-	+		-	-	
863645	45	m	+[1]		+[1]	-		+	+	
90262	49	m	+		-	+		-	+	
91092	54	f	-		-	+		-	-	
850050	56	f	-		-	+		+	+	
92140	62	m	-		-	+		+	+	
88252	66	f	-		-	-		-	-	
83170	70	f	+		-	+		+[1]	+	
93104	70	m	+		-	+		+	+	
91118	73	f	-		-	+		-	-	

	90349	77	m	-	-	+	+	-
	93099	77	m	-	-	-	+	+
	93225	81	m	-	-	-	+	+
	93101	83	f	-	-	-	+	+
	91045	85	f	-	-	+	+	-
	90345	86	m	-	-	+	-	+
	91061	88	m	+	-	-	-	+
	86004	90	f	-	-	-	+	+
	93105	92	f	-	-	+	+	+
	% pos. standing			26%	5%	63%	63%	63%

† also in glia (possible gliosis)

Table 3. Immunoreactivities in the human hippocampus for various neuronal proteins of which the mRNA is expressed in the +1 reading frame. Tissues were obtained from controls and neuropathologically confirmed Alzheimer and Down syndrome cases.

autopsy no.	age (years)	sex (m/f)	neuropathological state plaques	tangles	βAPP^{+1}	Tau^{+1}	Ubi-B^{+1}
N n-demented controls							
89004	34	m	-	-	-	-	-
81267	43	m	-	-	-	+	-
88149	58	m	-	+[a]	-	-	-
90183	65	f	-	-	-	-	-
90202	72	m	+[b]	+[c]	-	-	+
91090	80	f	+[b]	+[c]	-	-	+
91126	82	f	-	+[c]	-	-	+
90203	85	m	+[b]	+[b]	-	+	+
81033	90	f	+[b]	+[b]	-	-	+
90206	90	f	-	+[a]	-	-	+
% pos. staining					0%	20%	60%

TABLE 3 (CONTINUED)

Alzh imer cases								
89166	40	m	+[a]	+[c]	-	+	+	
863645	45	m	+[b]	+[c]	-	+	+	
90262	49	m	+[b]	+[c]	-	+	+	
91092	54	f	+[b]	+[c]	+	+	+	
850050	56	f	+[b]	+[a]	-	-	+	
92140	61	m	+[b]	+[c]	+	+	+	
88252	66	f	+[c]	+[c]	-	-	+	
83170	70	f	+[c]	+[c]	+	+	+	
93104	70	m	+[c]	+[c]	+	+	-	
91118	73	f	+[b]	+[c]	-	+	+	
90349	77	m	+[b]	+[c]	+	+	+	
93099	77	m	+[b]	+[c]	+	+	+	
93225	81	m	+[b]	+[c]	+	+	+	
93101	83	f	+[b]	+[c]	-	+	+	
88109	85	f	+[a]	+[b]	-	+	+	
91045	85	f	+[b]	+[c]	-	+	+	
90345	86	m	+[b]	+[c]	+	-	+	
91061	88	m	+[b]	+[c]	+	+	+	
86004	90	f	+[c]	+[c]	-	+	+	
93105	92	f	+[c]	+[c]	+	+	+	
% pos. staining					50%	85%	95%	
Down's syndrome								
25002	54	f	+[b]	+[c]	+	+	+	
92272	58	f	+[c]	+[c]	+	+	+	
89154	59	f	+[b]	+[c]	+	+	+	
25001	62	f	+[c]	+[c]	+	+	+	
95325	63	f	-	-	-	+	-	
94146	64	m	+[b]	+[c]	+	+	+	
93048	67	f	+[c]	+[c]	-	+	+	
% positive staining					71%	100%	86%	

* Number of plaques (all types) and tangles as revealed by Congo and Bodian silver staining: a) few, b) moderate, c) many.

Table 4. SUMMARY OF RESULTS

	Frontal cortex (area 11)			Temporal cortex (area 38)			Hippocampus		
	Amy^{+1}	Tau^{+1}	Ubi^{+1}	Amy^{+1}	Tau^{+1}	Ubi^{+1}	Amy^{+1}	Tau^{+1}	Ubi^{+1}
non dementing controls (n=10)	0	20	0	0	20	10*	0	20	60*
Alzheimer's disease (n=20)	15	80	80	40	100	95	50	85	95
Down syndrome (n=7)	86	100	86	86	100	86	71	100	86

Controls were sex- and age-matched.
* in old non-demented patients with some age related neuropathology (tangles, plaques)

One Down syndrome patient (#953251, age 63 years) does not show important neuropathology (no neurofibrillary tangles!). In this patient no Amy^{+1}, Ubi^{+1} and ALZ-50 (marker for early Alzheimer changes) immunoreactivity can be seen. Tau^{+1} immunoreactivity is present in cells resembling microglia. This case may be a Down patient with incomplete trisomy chromosome 21.

Amy^{+1} and Ubi^{+1} immunoreactivity are absent in the substantia nigra and striatum of 11 Parkinson patients, except for the striatum of one case (#90047) who also shows signs of Alzheimer's disease (tangles, plaques).

Amy^{+1} and Ubi^{+1} are absent from 1 patient (#93201) with frontal lobe dementia (Pick's disease), whereas Tau^{+1} immunoreactivity is present in microglia. Tau^{+1} immunoreactivity is not Alzheimer specific, since it also occurs in Parkinsons patients and even in the substantia nigra and striatum of age-matched controls. It is possibly a good marker for microglia.

Amy^{+1} and Ubi^{+1} immunoreactivity coexist in tangles and are both present in ALZ-50 positive neuropathological structures (e.g. tangles).

Ubi^{+1} immunoreactivity coexists in a subpopulation of wild-type Ubiquitin immunoreactive cells.

Ubi^{+1} immunoreactivity shows an age-dependant expression in the hippocampus of the control group (from 72 years onwards).

TABLE 4 (CONTINUED)

TABLE 5

	BASE PAIRS (CODING SEQUENCE OF LONGEST FORM)	GAGAG MOTIFS: EXPECTED NUMBER (1:1024)	GAGAG MOTIFS: ACTUAL NUMBER
βAPP	2234	2.2	7
TAU	1096	1.1	-
UBIQUITIN	687	0.7	2
APOLIPOPROTEIN E4	953	0.9	-
MAP2	5475	5.3	11
NF-LOW (68K)	582	0.6	3
NF - MEDIUM (145K)	2748	2.7	3
NF HIGH (200K)	3063	3.1	2

TABLE 6

	CHROMOSOME	EXON NO.	MOL. WEIGHT (kDa) LONGEST FORM	BASE PAIRS CODING SEQ. LONGEST FORM (total genomic seq.-kb)	GAGAG MOTIFS / EXPECTED NUMBER (1:1024)	GAGAG MOTIFS \ ACTUAL NUMBER	EXON(S) FROM WHICH +1 PEPTIDE IS DERIVED	PREDICTED MOL. WEIGHT OF +1 PEPTIDE	+1 PEPTIDE (C-terminus)
βAPP[1]	21q21.3-q22.05	18	135	2234 (170)	2.2	7	9/10	38	RGRTSSKELA
TAU[1]	17q21	15	67	1096 (100)	1.1	-	13	36	HGRLAPARHAS
UBIQUITIN[1]	17p11.1-p12	2	26	687 (1.8)	0.7	2	2	11	YADLREDPDRQ+ RQDHHPGSGAQ

References:
1. Yoshikai et al., Gene 87, 257, 1990; Selkoe et al., Proc. Natl. Acad. Sci USA 85, 7341, 1988; 2. Neve et al., J. Mol. Brain Res. 1, 271, 1986; Andreadis et al., Biochemistry, 31, 10626, 1992, in exon 4A of big Tau five GAGAG motifs are present; 3. Baker and Board, Nucl. Acids Res. 15, 443, 1987, Webb et al., Am. J. Hum. Genet. 46, 308, 1990.

Table 7

RGRTSSKELA
HGRLAPARHAS
YADLREDPDRQ
RQDHHPGSGAQ
GAPRLPPAQAA
KTRFQRKGPS
PGNRSMGHE
EAEGGSRS
VGAARDSRAA

CLAIMS

1. A method for the diagnosis of a disease caused by or associated with a gene having a somatic mutation giving rise to a frameshift mutation comprising:

i. providing a biological sample from a patient suspected of having or developing said disease; and

ii. detecting in said sample the presence of a mutant gene having a frameshift mutation or a protein encoded thereby,

wherein detection is indicative of the disease.

2. The method of claim 1, wherein the frameshift mutation comprises a deletion or an insertion of a nucleotide.

3. The method of claim 2, wherein the frameshift mutation is associated with the nucleotide sequence GAGA.

4. The method of claim 3, wherein the frameshift mutation comprises a dinucleotide mutation associated with a nucleotide sequence comprising GAGA.

5. The method of claim 3, wherein said sequence comprises GAGAX, where X is one of G, A, T, or C.

6. The method of claim 3, wherein said sequence comprises GAGAG or GAGAC.

7. The method of claim 1, wherein the disease is cancer or a neurodegenerative disease.

8. The method of claim 7, wherein the disease is Parkinson's disease, Alzheimer's disease, or Down's Syndrome.

9. The method of claim 1 wherein the gene having a frameshift mutation encodes the β amyloid precursor protein, the Tau protein, ubiquitin, apolipoprotein-E_4 (Apo-E_4),

microtubule associated protein II (MAP 2) or the neurofilament proteins.

10. The method of claim 1 wherein the biological sample comprises body fluid or tissue.

11. The method of claim 10 wherein said body fluid comprises cerebral spinal fluid or blood.

12. The method of claim 10, wherein the tissue comprises skin or nose epithelium.

13. The method of claim 1, wherein the mutant gene is detected by formation of a nucleic acid duplex wherein a first strand of said duplex comprises a nucleic acid probe having a sequence complementary to part of the mutant gene encompassing the mutation giving rise to the frameshift mutation, and the second strand of said duplex comprises a nucleic acid sequence of the mutant gene which is complementary to said probe.

14. The method of claim 1, wherein the mutant gene is detected using PCR to amplify a fragment of the mutant gene encompassing the mutation giving rise to the frameshift, and then probing for the amplified fragment using a nucleic acid probe having a sequence complementary to part of the mutant gene encompassing the mutation giving rise to the frameshift mutation, or by sequencing the amplified fragment.

15. The method of claim 1, wherein the protein encoded by the mutant gene is detected using an antibody molecule having specificity for the mutant protein and not for the wild-type protein.

16. A method for identifying diseases caused by or associated with a gene having a somatic mutation giving rise to a frameshift mutation comprising:

i. providing the sequence of a gene suspected of

being involved in the pathogenesis of a disease;

ii. identifying the sequence of the mutant protein encoded by the gene sequence following a frameshift mutation;

iii. preparing a probe to the mutant protein or a fragment thereof; and

iv. probing a biological sample from a patient having the disease and a biological sample from a patient not having the disease,

wherein the presence of said mutant protein in a biological sample from a patient having the disease and the absence of said mutant protein in a biological sample from a patient not having the disease indicates that the presence of the mutant protein in a biological sample is a marker for the disease or susceptibility to the disease.

17. A diagnostic kit for diagnosing a disease caused by or associated with a gene having a somatic mutation giving rise to a frameshift mutation, the kit comprising:

i. a labelled nucleic acid probe having a sequence complementary to part of the mutant gene which encompasses the mutation which leads to the frameshift mutation; and

ii. packaging materials therefor.

18. A diagnostic kit for diagnosing a disease caused by or associated with a gene having a somatic mutation giving rise to a frameshift mutation comprising:

i. a pair of primers for use in a PCR reaction, wherein said pair comprises sequences complementary to sequences on either side of the mutation which gives rise to the frameshift mutation, and reagents necessary for performing a PCR reaction; and

ii. packaging materials therefor.

19. A diagnostic kit for diagnosing a disease caused by or associated with at least one gene having one or more somatic mutations giving rise to a frameshift mutation

comprising:

i. an antibody molecule having specificity for the mutant protein and not the wild-type protein; and

ii. packaging materials therefor.

20. A recombinant gene having a frameshift mutation, as described in of any one of claims 1 to 9.

21. The gene of claim 20 encoding at least part of the protein sequence designated +1 or +2 shown in any one of Figures 2 to 9.

22. A mutant protein encoded by the gene of claim 20 or 21.

23. An immunogenic fragment of the mutant protein of claim 22.

24. The mutant protein of claim 22 or the immunogenic fragment of claim 23, comprising the amino acid sequence:

RGRTSSKELA;
HGRLAPARHAS;
YADLREDPDRQ;
RQDHHPGSGAQ;
GAPRLPPAQAA;
KTRFQRKGPS;
PGNRSMGHE;
EAEGGSRS; or
VGAARDSRAA.

25. A pharmaceutical composition comprising a DNA repair enzyme admixed with a pharmaceutically acceptable carrier.

26. A method of treatment and/or prevention of a disease caused by or associated with a gene having a somatic mutation giving rise to a frameshift mutation, comprising administering a vector comprising an expressible gene

encoding a repair enzyme to a patient suffering from or susceptible to the disease.

27. A vector comprising an expressible gene encoding a repair enzyme.

28. A method of treatment and/or prevention of a disease caused by or associated with a gene having a somatic mutation giving rise to a frameshift mutation, comprising administering a vector comprising an expressible gene encoding a ribozyme to a patient suffering from or susceptible to the disease.

29. A vector comprising an expressible gene encoding a ribozyme.

30. A pharmaceutical composition comprising the wild-type analog of a mutant protein in admixture with a pharmaceutically acceptable carrier.

31. A method of treatment and/or prevention of a disease caused by or associated with a gene having a somatic mutation giving rise to a frameshift mutation, comprising administering a vector comprising an expressible gene encoding the wild-type version of the mutated gene to a patient suffering from or susceptible to the disease.

32. A vector comprising an expressible gene encoding the wild-type version of a mutated gene.

33. A host cell containing a recombinant gene comprising a frameshift mutation, as described in claim 20 or 21.

34. A transgenic animal having somatic cells which contain a transgene comprising a frameshift mutation, as described in claim 20 or 21.

35. The use of a vector encoding one or more repair enzymes under the control of a promoter in therapy.

36. The use of a vector encoding one or more repair enzymes under the control of a promoter in the manufacture of a composition for the treatment of a disease caused by or associated with at least one gene having one or more somatic mutations giving rise to a frameshift mutation.

37. The use of a vector encoding a ribozyme under the control of a promoter in therapy.

38. The use of a vector encoding a ribozyme under the control of a promoter in the manufacture of a composition for the treatment of a disease caused by or associated with at least one gene having one or more somatic mutations giving rise to a frameshift mutation.

39. The use of a vector encoding the wildtype version of the mutated gene under the control of a promoter in therapy.

40. The use of a vector encoding the wildtype version of the mutated gene under the control of a promoter in the manufacture of a composition for treating a disease caused by or associated with at least one gene having one or more somatic mutations giving rise to a frameshift mutation.

41. The use of more than one of the vectors of claim 27, 29 and 32 in any combination in therapy.

42. The use of more than one of the vectors of claim 27, 29 and 32 in any combination in the treatment and/or prevention of a disease caused by or associated with at least one gene having one or more somatic mutations giving rise to a frameshift mutation.

FIG. 1(A)

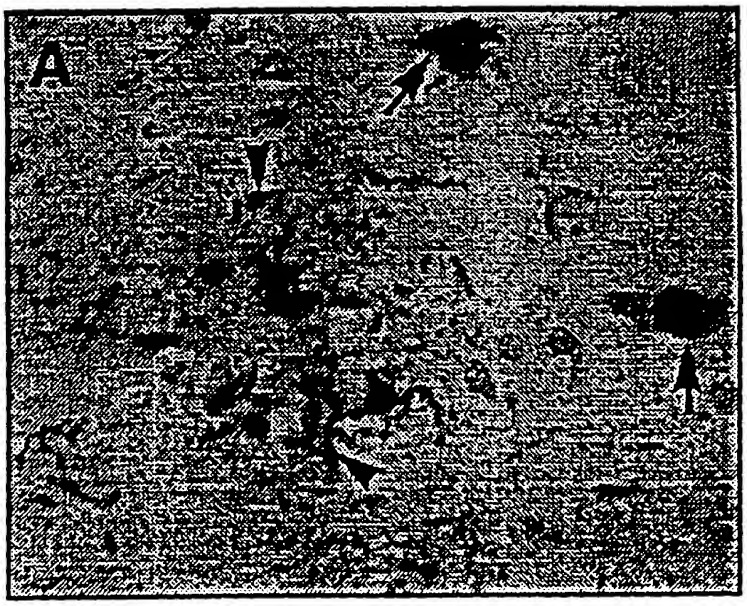
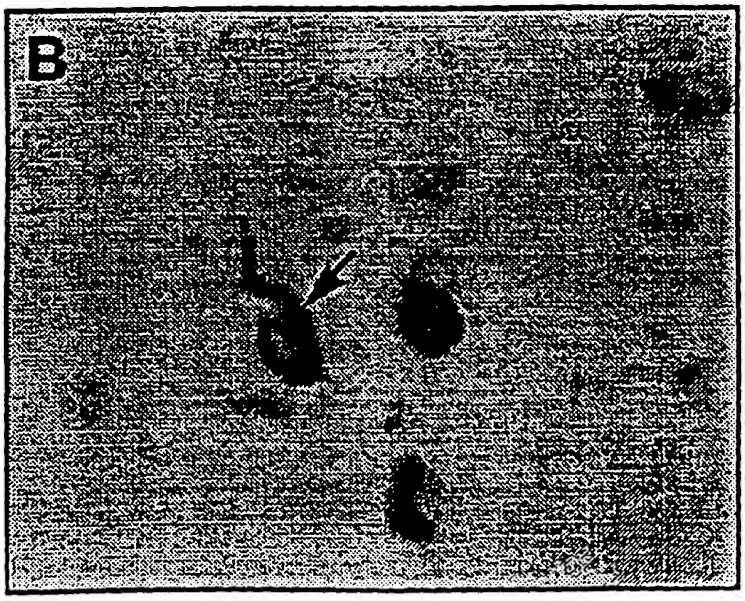


FIG. 1(B)



(Linear) MAP of: Freda4.Seq check: 6510 from: 147 to: 2300

LOCUS HUMAFPA4 3353 bp ss-mRNA PRI 15-JUN-1989
DEFINITION Human amyloid A4 mRNA, complete cds.
ACCESSION Y00264
KEYWORDS amyloid fibril protein; cell surface glycoprotein.
SOURCE human (Homo sapiens).
ORGANISM Homo sapiens . . .

FIG. 2

With 1 enzymes: NOTI

September 14, 1993 11:31 ..

```
       ATGCTGCCCGGTTTGGCACTGCTCCTGCTGGCCGCCTGGACGGCTCGGGCGCTGGAGGTA
   147 ---+---------+---------+---------+---------+---------+------ 206
       TACGACGGGCCAAACCGTGACGAGGACGACCGGCGGACCTGCCGAGCCCGCGACCTCCAT

a       C  C  P  V  W  H  C  S  C  W  P  P  G  R  L  G  R  W  R  Y  - +2≡(-1)
b        A  A  R  F  G  T  A  P  A  G  R  L  D  G  S  G  A  G  G  T - +1≡(-2)
c      M  L  P  G  L  A  L  L  L  L  A  A  W  T  A  R  A  L  E  V  - wt

       CCCACTGATGGTAATGCTGGCCTGCTGGCTGAACCCCAGATTGCCATGTTCTGTGGCAGA
   207 ---+---------+---------+---------+---------+---------+------ 266
       GGGTGACTACCATTACGACCGGACGACCGACTTGGGGTCTAACGGTACAAGACACCGTCT

a       P  L  M  V  M  L  A  C  W  L  N  P  R  L  P  C  S  V  A  D  -
b        H  *  W  *  C  W  P  A  G  *  T  P  D  C  H  V  L  W  Q  T -
c      P  T  D  G  N  A  G  L  L  A  E  P  Q  I  A  M  F  C  G  R  -

       CTGAACATGCACATGAATGTCCAGAATGGGAAGTGGGATTCAGATCCATCAGGGACCAAA
   267 ---+---------+---------+---------+---------+---------+------ 326
       GACTTGTACGTGTACTTACAGGTCTTACCCTTCACCCTAAGTCTAGGTAGTCCCTGGTTT

a       *  T  C  T  *  M  S  R  M  G  S  G  I  Q  I  H  Q  G  P  K  -
b        E  H  A  H  E  C  P  E  W  E  V  G  F  R  S  I  R  D  Q  N -
c      L  N  M  H  M  N  V  Q  N  G  K  W  D  S  D  P  S  G  T  K  -
```

```
             ACCTGCATTGATACCAAGGAAGGCATCCTGCAGTATTGCCAAGAAGTCTACCCTGAACTG
        327  ---+---------+---------+---------+---------+---------+------ 386
             TGGACGTAACTATGGTTCCTTCCGTAGGACGTCATAACGGTTCTTCAGATGGGACTTGAC

a             P  A  L  I  P  R  K  A  S  C  S  I  A  K  K  S  T  L  N  C  -
b              L  H  *  Y  Q  G  R  H  P  A  V  L  P  R  S  L  P  *  T  A -
c            T  C  I  D  T  K  E  G  I  L  Q  Y  C  Q  E  V  Y  P  E  L   -

             CAGATCACCAATGTGGTAGAAGCCAACCAACCAGTGACCATCCAGAACTGGTGCAAGCGG
        387  ---+---------+---------+---------+---------+---------+------ 446
             GTCTAGTGGTTACACCATCTTCGGTTGGTTGGTCACTGGTAGGTCTTGACCACGTTCGCC

a             R  S  P  M  W  *  K  P  T  N  Q  *  P  S  R  T  G  A  S  G  -
b              D  H  Q  C  G  R  S  Q  P  T  S  D  H  P  E  L  V  Q  A  G -
c            Q  I  T  N  V  V  E  A  N  Q  P  V  T  I  Q  N  W  C  K  R   -

             GGCCGCAAGCAGTGCAAGACCCATCCCCACTTTGTGATTCCCTACCGCTGCTTAGTTGGT
        447  ---+---------+---------+---------+---------+---------+------ 506
             CCGGCGTTCGTCACGTTCTGGGTAGGGGTGAAACACTAAGGGATGGCGACGAATCAACCA

a             A  A  S  S  A  R  P  I  P  T  L  *  F  P  T  A  A  *  L  V  -
b              P  Q  A  V  Q  D  P  S  P  L  C  D  S  L  P  L  L  S  W  * -
c            G  R  K  Q  C  K  T  H  P  H  F  V  I  P  Y  R  C  L  V  G   -

             GAGTTTGTAAGTGATGCCCTTCTCGTTCCTGACAAGTGCAAATTCTTACACCAG[GAGAG]G
        507  ---+---------+---------+---------+---------+---------+------ 566
             CTCAAACATTCACTACGGGAAGAGCAAGGACTGTTCACGTTTAAGAATGTGGTCCTCTCC

a             S  L  *  V  M  P  F  S  F  L  T  S  A  N  S  Y  T  R  R  G  -
```

FIG. 2(I)

FIG. 2(II)

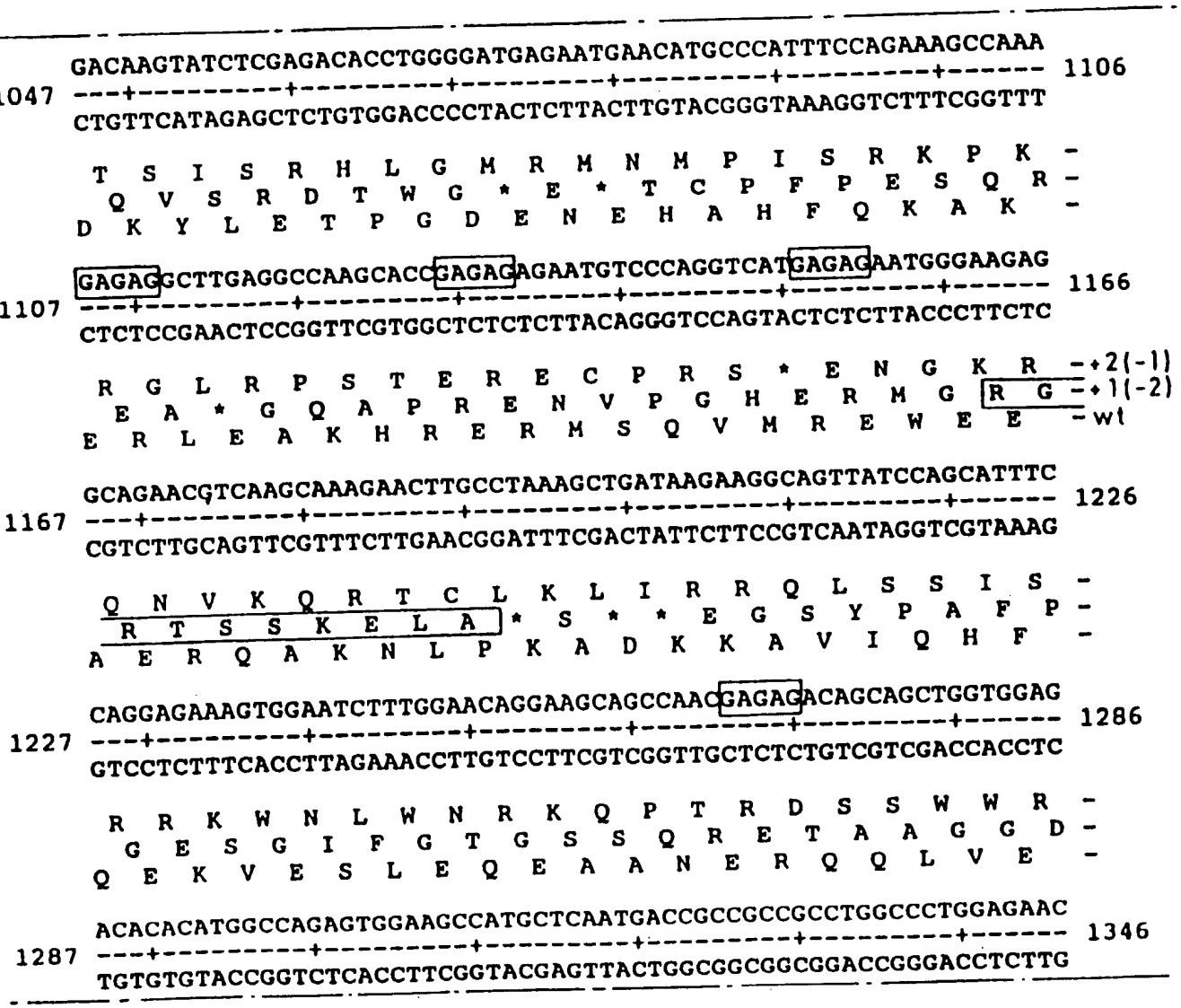

GACAAGTATCTCGAGACACCTGGGGATGAGAATGAACATGCCCATTTCCAGAAAGCCAAA
1047
1106
CTGTTCATAGAGCTCTGTGGACCCCTACTCTTACTTGTACGGGTAAAGGTCTTTCGGTTT
T S I S R H L G M R M N M P I S R K P K
Q V S R D T W G * E * T C P F P E S Q R
D K Y L E T P G D E N E H A H F Q K A K
GAGAGGCTTGAGGCCAAGCACCGAGAGAGAATGTCCCAGGTCATGAGAGAATGGGAAGAG
1107
1166
CTCTCCGAACTCCGGTTCGTGGCTCTCTCTTACAGGGTCCAGTACTCTCTTACCCTTCTC
R G L R P S T E R E C P R S * E N G K R
E A * G Q A P R E N V P G H E R M G R G
E R L E A K H R E R M S Q V M R E W E E
+2(-1)
+1(-2)
wt
GCAGAACGTCAAGCAAAGAACTTGCCTAAAGCTGATAAGAAGGCAGTTATCCAGCATTTC
1167
1226
CGTCTTGCAGTTCGTTTCTTGAACGGATTTCGACTATTCTTCCGTCAATAGGTCGTAAAG
Q N V K Q R T C L K L I R R Q L S S I S
R T S S K E L A * S * * E G S Y P A F P
A E R Q A K N L P K A D K K A V I Q H F
CAGGAGAAAGTGGAATCTTTGGAACAGGAAGCAGCCAACGAGAGACAGCAGCTGGTGGAG
1227
1286
GTCCTCTTTCACCTTAGAAACCTTGTCCTTCGTCGGTTGCTCTCTGTCGTCGACCACCTC
R R K W N L W N R K Q P T R D S S W W R
G E S G I F G T G S S Q R E T A A G G D
Q E K V E S L E Q E A A N E R Q Q L V E
ACACACATGGCCAGAGTGGAAGCCATGCTCAATGACCGCCGCCGCCTGGCCCTGGAGAAC
1287
1346
TGTGTGTACCGGTCTCACCTTCGGTACGAGTTACTGGCGGCGGCGGACCGGGACCTCTTG

FIG. 2(III)

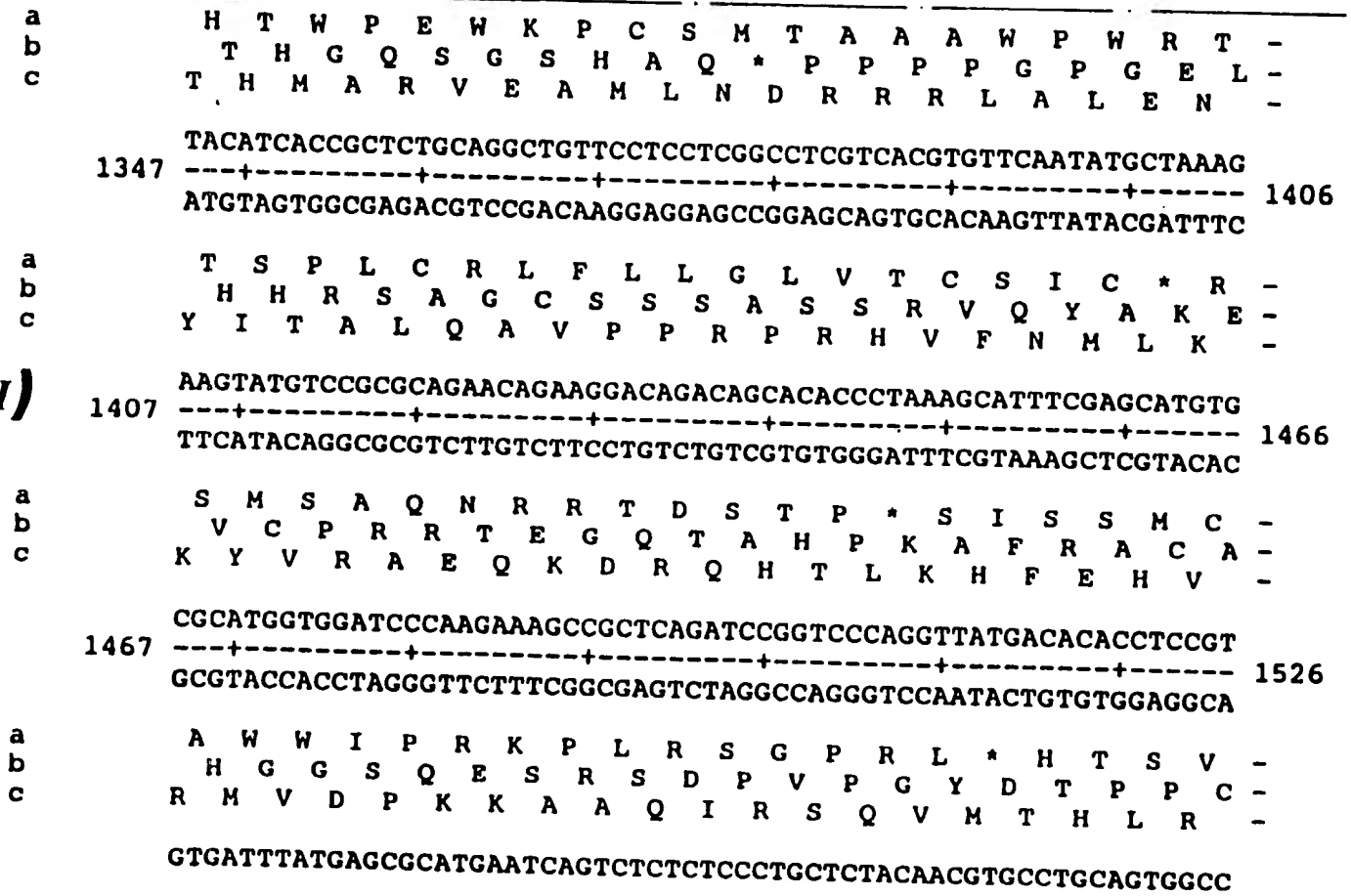

a H T W P E W K P C S M T A A A W P W R T -
b T H G Q S G S H A Q * P P P P G P G E L -
c T H M A R V E A M L N D R R R L A L E N -
TACATCACCGCTCTGCAGGCTGTTCCTCCTCGGCCTCGTCACGTGTTCAATATGCTAAAG
1347 ---+---------+---------+---------+---------+---------+---------+------ 1406
ATGTAGTGGCGAGACGTCCGACAAGGAGGAGCCGGAGCAGTGCACAAGTTATACGATTTC
a T S P L C R L F L L G L V T C S I C * R -
b H H R S A G C S S S A S S R V Q Y A K E -
c Y I T A L Q A V P P R P R H V F N M L K -
AAGTATGTCCGCGCAGAACAGAAGGACAGACAGCACACCCTAAAGCATTTCGAGCATGTG
1407 ---+---------+---------+---------+---------+---------+---------+------ 1466
TTCATACAGGCGCGTCTTGTCTTCCTGTCTGTCGTGTGGGATTTCGTAAAGCTCGTACAC
a S M S A Q N R R T D S T P * S I S S M C -
b V C P R R T E G Q T A H P K A F R A C A -
c K Y V R A E Q K D R Q H T L K H F E H V -
CGCATGGTGGATCCCAAGAAAGCCGCTCAGATCCGGTCCCAGGTTATGACACACCTCCGT
1467 ---+---------+---------+---------+---------+---------+---------+------ 1526
GCGTACCACCTAGGGTTCTTTCGGCGAGTCTAGGCCAGGGTCCAATACTGTGTGGAGGCA
a A W W I P R K P L R S G P R L * H T S V -
b H G G S Q E S R S D P V P G Y D T P P C -
c R M V D P K K A A Q I R S Q V M T H L R -
GTGATTTATGAGCGCATGAATCAGTCTCTCTCCCTGCTCTACAACGTGCCTGCAGTGGCC

FIG. 3

```
(Linear) MAP of: Fredtau.Seq  check: 9711  from: 38  to: 1096

RL;HSTAUA      - Human microtubule-associated protein tau mRNA, complete cds
ID   HSTAUA      standard; RNA; PRI; 1108 BP.
XX
AC   J03778;
XX
DT   04-OCT-1988 (Rel. 17, Created) . . .

 With 1 enzymes: NOTI

                                September 14, 1993  12:12  ..

      ATGGCTGAGCCCCGCCAGGAGTTCGAAGTGATGGAAGATCACGCTGGGACGTACGGGTTG
   38 --+---------+---------+---------+---------+---------+------- 97
      TACCGACTCGGGGCGGTCCTCAAGCTTCACTACCTTCTAGTGCGACCCTGCATGCCCAAC

a       G  *  A  P  P  G  V  R  S  D  G  R  S  R  W  D  V  R  V  G  --2(+1)
b     M  A  E  P  R  Q  E  F  E  V  M  E  D  H  A  G  T  Y  G  L    - 0
c      W  L  S  P  A  R  S  S  K  *  W  K  I  T  L  G  R  T  G  W  --1(+2)

      GGGGACAGGAAAGATCAGGGGGGCTACACCATGCACCAAGACCAAGAGGGTGACACGGAC
   98 --+---------+---------+---------+---------+---------+------- 157
      CCCCTGTCCTTTCTAGTCCCCCCGATGTGGTACGTGGTTCTGGTTCTCCCACTGTGCCTG

a       G  Q  E  R  S  G  G  L  H  H  A  P  R  P  R  G  *  H  G  R  -
b     G  D  R  K  D  Q  G  G  Y  T  M  H  Q  D  Q  E  G  D  T  D    -
c      G  T  G  K  I  R  G  A  T  P  C  T  K  T  K  R  V  T  R  T  -

      GCTGGCCTGAAAGCTGAAGAAGCAGGCATTGGAGACACCCCCAGCCTGGAAGACGAAGCT
  158 --+---------+---------+---------+---------+---------+------- 217
      CGACCGGACTTTCGACTTCTTCGTCCGTAACCTCTGTGGGGGTCGGACCTTCTGCTTCGA

a       W  P  E  S  *  R  S  R  H  W  R  H  P  Q  P  G  R  R  S  C  -
b     A  G  L  K  A  E  E  A  G  I  G  D  T  P  S  L  E  D  E  A    -
c      L  A  *  K  L  K  K  Q  A  L  E  T  P  P  A  W  K  T  K  L  -
```

```
      GCTGGTCACGTGACCCAAGCTCGCATGGTCAGTAAAAGCAAAGACGGGACTGGAAGCGAT
  218 ---+---------+---------+---------+---------+---------+------- 277
      CGACCAGTGCACTGGGTTCGAGCGTACCAGTCATTTTCGTTTCTGCCCTGACCTTCGCTA

a       W  S  R  D  P  S  S  H  G  Q  *  K  Q  R  R  D  W  K  R  *  -
b     A  G  H  V  T  Q  A  R  M  V  S  K  S  K  D  G  T  G  S  D   -
c      L  V  T  *  P  K  L  A  W  S  V  K  A  K  T  G  L  E  A  M  -

      GACAAAAAAGCCAAGGGGGCTGATGGTAAAACGAAGATCGCCACACCGCGGGGAGCAGCC
  278 ---+---------+---------+---------+---------+---------+------- 337
      CTGTTTTTTCGGTTCCCCCGACTACCATTTTGCTTCTAGCGGTGTGGCGCCCCTCGTCGG

a       Q  K  S  Q  G  G  *  W  *  N  E  D  R  H  T  A  G  S  S  P  -
b     D  K  K  A  K  G  A  D  G  K  T  K  I  A  T  P  R  G  A  A   -  100
c      T  K  K  P  R  G  L  M  V  K  R  R  S  P  H  R  G  E  Q  P  -

      CCTCCAGGCCAGAAGGGCCAGGCCAACGCCACCAGGATTCCAGCAAAAACCCCGCCCGCT
  338 ---+---------+---------+---------+---------+---------+------- 397
      GGAGGTCCGGTCTTCCCGGTCCGGTTGCGGTGGTCCTAAGGTCGTTTTTGGGGCGGGCGA

a       S  R  P  E  G  P  G  Q  R  H  Q  D  S  S  K  N  P  A  R  S  -
b     P  P  G  Q  K  G  Q  A  N  A  T  R  I  P  A  K  T  P  P  A   -
c      L  Q  A  R  R  A  R  P  T  P  P  G  F  Q  Q  K  P  R  P  L  -

      CCAAAGACACCACCCAGCTCTGGTGAACCTCCAAAATCAGGGGATCGCAGCGGCTACAGC
  398 ---+---------+---------+---------+---------+---------+------- 457
      GGTTTCTGTGGTGGGTCGAGACCACTTGGAGGTTTTAGTCCCCTAGCGTCGCCGATGTCG

a       K  D  T  T  Q  L  W  *  T  S  K  I  R  G  S  Q  R  L  Q  Q  -
b     P  K  T  P  P  S  S  G  E  P  P  K  S  G  D  R  S  G  Y  S   -
```

FIG. 3(I)

FIG. 3(II)

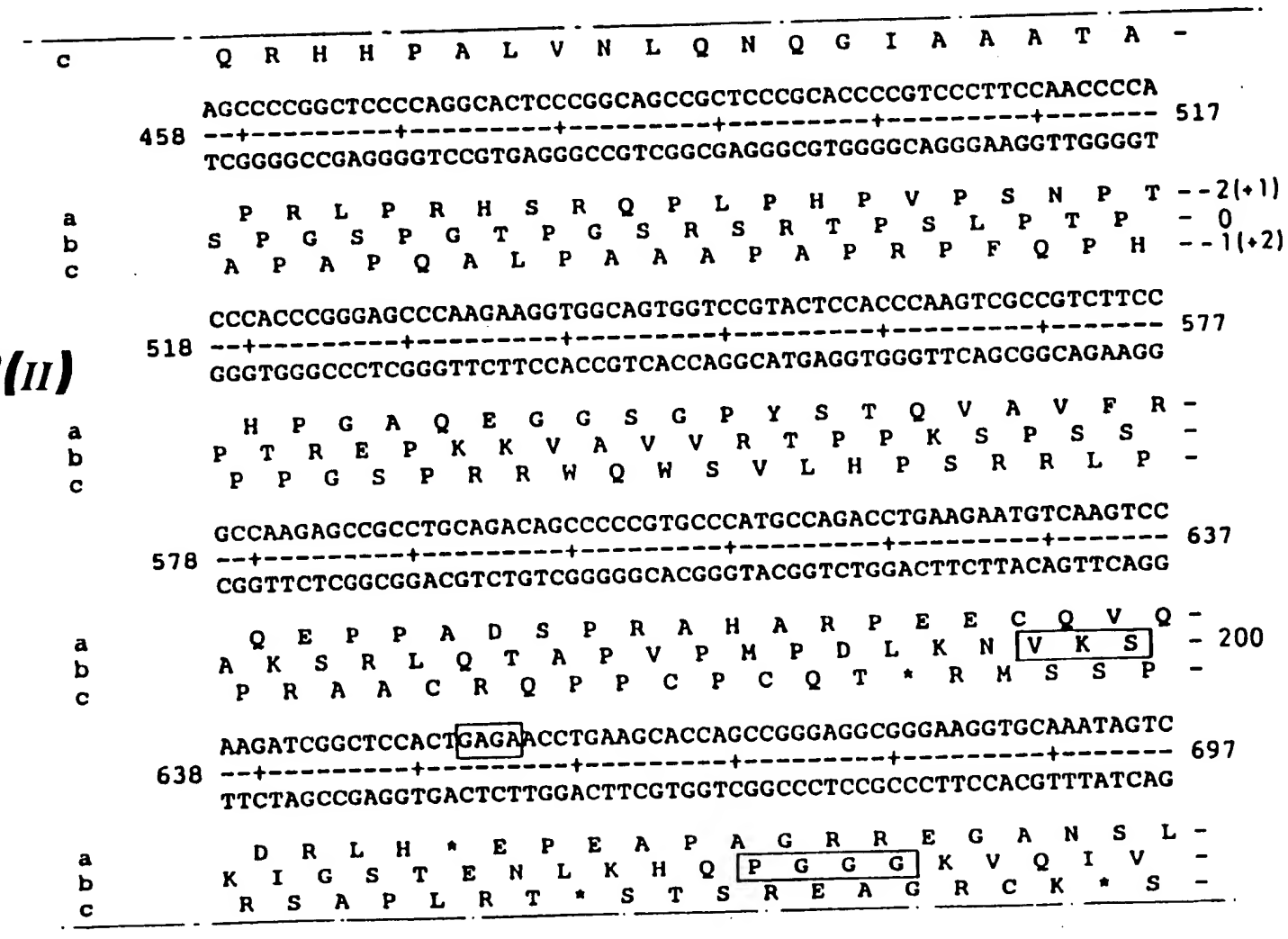
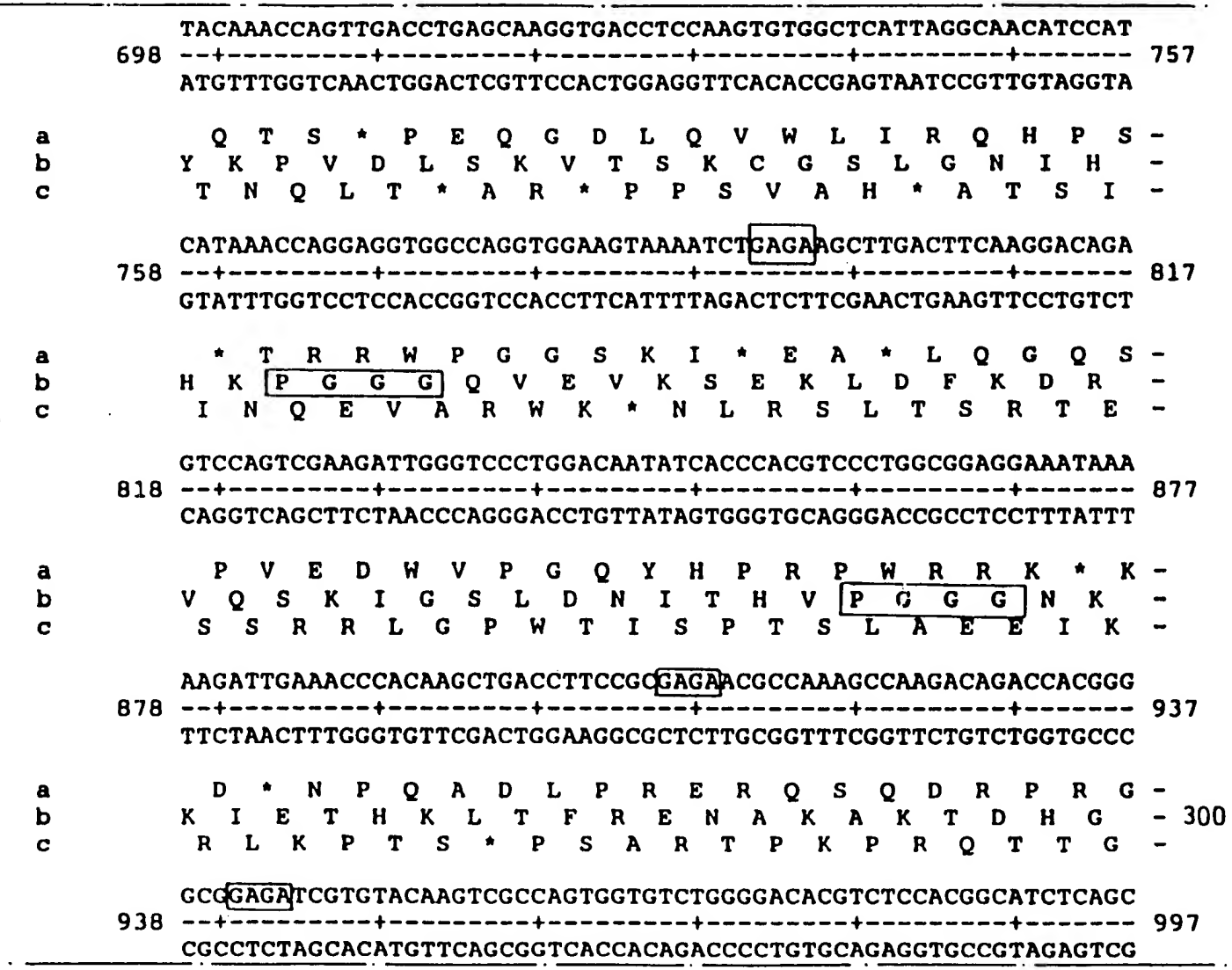

c Q R H H P A L V N L Q N Q G I A A A T A -
AGCCCCGGCTCCCCAGGCACTCCCGGCAGCCGCTCCCGCACCCCGTCCCTTCCAACCCCA
458 517
TCGGGGCCGAGGGGTCCGTGAGGGCCGTCGGCGAGGGCGTGGGGCAGGGAAGGTTGGGGT
a P R L P R H S R Q P L P H P V P S N P T --2(+1)
b S P G S P G T P G S R S R T P S L P T P - 0
c A P A P Q A L P A A A P A P R P F Q P H --1(+2)
CCCACCCGGGAGCCCAAGAAGGTGGCAGTGGTCCGTACTCCACCCAAGTCGCCGTCTTCC
518 577
GGGTGGGCCCTCGGGTTCTTCCACCGTCACCAGGCATGAGGTGGGTTCAGCGGCAGAAGG
a H P G A Q E G G S G P Y S T Q V A V F R -
b P T R E P K K V A V V R T P P K S P S S S -
c P P G S P R R W Q W S V L H P S R R L P -
GCCAAGAGCCGCCTGCAGACAGCCCCCGTGCCCATGCCAGACCTGAAGAATGTCAAGTCC
578 637
CGGTTCTCGGCGGACGTCTGTCGGGGGCACGGGTACGGTCTGGACTTCTTACAGTTCAGG
a Q E P P A D S P R A H A R P E E C Q V Q -
b A K S R L Q T A P V P M P D L K N V K S - 200
c P R A A C R Q P P C P C Q T * R M S S P -
AAGATCGGCTCCACTGAGAACCTGAAGCACCAGCCGGGAGGCGGGAAGGTGCAAATAGTC
638 697
TTCTAGCCGAGGTGACTCTTGGACTTCGTGGTCGGCCCTCCGCCCTTCCACGTTTATCAG
a D R L H * E P E A P A G R R E G A N S L -
b K I G S T E N L K H Q P G G G K V Q I V -
c R S A P L R T * S T S R E A G R C K * S -

FIG. 3(III)


TACAAACCAGTTGACCTGAGCAAGGTGACCTCCAAGTGTGGCTCATTAGGCAACATCCAT
698 757
ATGTTTGGTCAACTGGACTCGTTCCACTGGAGGTTCACACCGAGTAATCCGTTGTAGGTA
a Q T S * P E Q G D L Q V W L I R Q H P S -
b Y K P V D L S K V T S K C G S L G N I H -
c T N Q L T * A R * P P S V A H * A T S I -
CATAAACCAGGAGGTGGCCAGGTGGAAGTAAAATCTGAGAAGCTTGACTTCAAGGACAGA
758 817
GTATTTGGTCCTCCACCGGTCCACCTTCATTTTAGACTCTTCGAACTGAAGTTCCTGTCT
a * T R R W P G G S K I * E A * L Q G Q S -
b H K P G G G Q V E V K S E K L D F K D R -
c I N Q E V A R W K * N L R S L T S R T E -
GTCCAGTCGAAGATTGGGTCCCTGGACAATATCACCCACGTCCCTGGCGGAGGAAATAAA
818 877
CAGGTCAGCTTCTAACCCAGGGACCTGTTATAGTGGGTGCAGGGACCGCCTCCTTTATTT
a P V E D W V P G Q Y H P R P W R R K * K -
b V Q S K I G S L D N I T H V P G G G N K -
c S S R R L G P W T I S P T S L A E E I K -
AAGATTGAAACCCACAAGCTGACCTTCCGCGAGAACGCCAAAGCCAAGACAGACCACGGG
878 937
TTCTAACTTTGGGTGTTCGACTGGAAGGCGCTCTTGCGGTTTCGGTTCTGTCTGGTGCCC
a D * N P Q A D L P R E R Q S Q D R P R G -
b K I E T H K L T F R E N A K A K T D H G - 300
c R L K P T S * P S A R T P K P R Q T T G -
GCGGAGATCGTGTACAAGTCGCCAGTGGTGTCTGGGGACACGTCTCCACGGCATCTCAGC
938 997
CGCCTCTAGCACATGTTCAGCGGTCACCACAGACCCCTGTGCAGAGGTGCCGTAGAGTCG

```
a          G  D  R  V  Q  V  A  S  G  V  W  G  H  V  S  T  A  S  Q  Q  - -2(+1)
b       A  E  I  V  Y  K  S  P  V  V  S  G  D  T  S  P  R  H  L  S     - 0
c        R  R  S  C  T  S  R  Q  W  C  L  G  T  R  L  H  G  I  S  A  - -1(+2)

        AATGTCTCCTCCACCGGCAGCATCGACATGGTAGACTCGCCCCAGCTCGCCACGCTAGCT
    998 --+---------+---------+---------+---------+---------+------- 1057
        TTACAGAGGAGGTGGCCGTCGTAGCTGTACCATCTGAGCGGGGTCGAGCGGTGCGATCGA

a          C  L  L  H  R  Q  H  R  H  G  R  L  A  P  A  R  H  A  S  * -
b       N  V  S  S  T  G  S  I  D  M  V  D  S  P  Q  L  A  T  L  A    -
c        M  S  P  P  P  A  A  S  T  W  *  T  R  P  S  S  P  R  *  L  -

        GACGAGGTGTCTGCCTCCCTGGCCAAGCAGGGTTTGTGA
   1058 --+---------+---------+---------+------ 1096
        CTGCTCCACAGACGGAGGGACCGGTTCGTCCCAAACACT

a          R  G  V  C  L  P  G  Q  A  G  F  V     -
b       D  E  V  S  A  S  L  A  K  Q  G  L  (*)   -
c        T  R  C  L  P  P  W  P  S  R  V  C       -

Enzymes that do cut:

  NONE

Enzymes that do not cut:

    NotI
```

FIG. 3(IV)

FIG. 4

(Linear) MAP of: Fredubi.Seq check: 2987 from: 1094 to: 1800

LOCUS HUMYUBG1 2118 bp ds-DNA PRI 15-MAR-1988
DEFINITION Human ubiquitin gene (3 repeats of 76aa)
ACCESSION X04803
KEYWORDS ubiquitin.
SOURCE human (Homo sapiens).
ORGANISM Homo sapiens . . .

With 1 enzymes: NOTI

September 14, 1993 11:58 ..

```
      ATGCAGATCTTCGTGAAAACCCTTACCGGCAAGACCATCACCCTTGAGGTGGAGCCCAGT
1094  ------+---------+---------+---------+---------+---------+--- 1153
      TACGTCTAGAAGCACTTTTGGGAATGGCCGTTCTGGTAGTGGGAACTCCACCTCGGGTCA

a       A  D  L  R  E  N  P  Y  R  Q  D  H  H  P  *  G  G  A  Q  *  - +1(-2)
b     M  Q  I  F  V  K  T  L  T  G  K  T  I  T  L  E  V  E  P  S   - wt
c      C  R  S  S  *  K  P  L  P  A  R  P  S  P  L  R  W  S  P  V  - +2(-1)

      GACACCATCGAAAATGTGAAGGCCAAGATCCAGGATAAGGAAGGCATTCCCCCCGACCAG
1154  ------+---------+---------+---------+---------+---------+--- 1213
      CTGTGGTAGCTTTTACACTTCCGGTTCTAGGTCCTATTCCTTCCGTAAGGGGGGCTGGTC

a       H  H  R  K  C  E  G  Q  D  P  G  *  G  R  H  S  P  R  P  A  -
b     D  T  I  E  N  V  K  A  K  I  Q  D  K  E  G  I  P  P  D  Q   -
c      T  P  S  K  M  *  R  P  R  S  R  I  R  K  A  F  P  P  T  S  -

      CAGAGGCTCATCTTTGCAGGCAAGCAGCTGGAAGATGGCCGTACTCTTTCTGACTACAAC
1214  ------+---------+---------+---------+---------+---------+--- 1273
      GTCTCCGAGTAGAAACGTCCGTTCGTCGACCTTCTACCGGCATGAGAAAGACTGATGTTG
```


WO 97/12992

PCT/IB96/01106

11/38

FIG. 4(1)

```
a          E  A  H  L  C  R  Q  A  A  G  R  W  P  Y  S  F  *  L  Q  H  -
b       Q  R  L  I  F  A  G  K  Q  L  E  D  G  R  T  L  S  D  Y  N     -
c        R  G  S  S  L  Q  A  S  S  W  K  M  A  V  L  F  L  T  T  T  -

                                                                        1
        ATCCAGAAGGAGTCGACCCTGCACCTGGTCCTGCGTCTGAGAGGTGGTATGCAGATCTTC
1274    ------+---------+---------+---------+---------+---------+---  1333
        TAGGTCTTCCTCAGCTGGGACGTGGACCAGGACGCAGACTCTCCACCATACGTCTAGAAG
                                                  UBI 1•1
a          P  E  G  V  D  P  A  P  G  P  A  S  E  R  W  Y  A  D  L  R  -
b       I  Q  K  E  S  T  L  H  L  V  L  R  L  R  G  G  M  Q  I  F     - REPEAT
c        S  R  R  S  R  P  C  T  W  S  C  V  *  E  V  V  C  R  S  S  -    2

        GTGAAGACCCTGACCGGCAAGACCATCACCCTGGAAGTGGAGCCCAGTGACACCATCGAA
1334    ------+---------+---------+---------+---------+---------+---  1393
        CACTTCTGGGACTGGCCGTTCTGGTAGTGGGACCTTCACCTCGGGTCACTGTGGTAGCTT
                                                  UBI 2•1
a          E  D  P  D  R  Q  D  H  H  P  G  S  G  A  Q  *  H  H  R  K  -
b       V  K  T  L  T  G  K  T  I  T  L  E  V  E  P  S  D  T  I  E     -
c        *  R  P  *  P  A  R  P  S  P  W  K  W  S  P  V  T  P  S  K  -

        AATGTGAAGGCCAAGATCCAGGATAAAGAAGGCATCCCTCCCGACCAGCAGAGGCTCATC
1394    ------+---------+---------+---------+---------+---------+---  1453
        TTACACTTCCGGTTCTAGGTCCTATTTCTTCCGTAGGGAGGGCTGGTCGTCTCCGAGTAG

a          C  E  G  Q  D  P  G  *  R  R  H  P  S  R  P  A  E  A  H  L  -
b       N  V  K  A  K  I  Q  D  K  E  G  I  P  P  D  Q  Q  R  L  I     -
c        M  *  R  P  R  S  R  I  K  K  A  S  L  P  T  S  R  G  S  S  -

        TTTGCAGGCAAGCAGCTGGAAGATGGCCGCACTCTTTCTGACTACAACATCCAGAAGGAG
1454    ------+---------+---------+---------+---------+---------+---  1513
        AAACGTCCGTTCGTCGACCTTCTACCGGCGTGAGAAAGACTGATGTTGTAGGTCTTCCTC

a          C  R  Q  A  A  G  R  W  P  H  S  F  *  L  Q  H  P  E  G  V  -
b       F  A  G  K  Q  L  E  D  G  R  T  L  S  D  Y  N  I  Q  K  E     -
```

FIG. 4(II)

```
c          L  Q  A  S  S  W  K  M  A  A  L  F  L  T  T  T  S  R  R  S  -

           TCGACCCTGCACCTGGTCCTGCGTCTGAGAGGTGGTATGCAGATCTTCGTGAAGACCCTG
1514       ------+---------+---------+---------+---------+---------+---  1573
           AGCTGGGACGTGGACCAGGACGCAGACTCTCCACCATACGTCTAGAAGCACTTCTGGGAC

a            D  P  A  P  G  P  A  S  E  R  W  Y  A  D  L  R  E  D  P  D  -+1(-2)
b          S  T  L  H  L  V  L  R  L  R  G  G  M  Q  I  F  V  K  T  L    -wt
c            R  P  C  T  W  S  C  V  *  E  V  V  C  R  S  S  *  R  P  *  -+2

           ACCGGCAAGACCATCACTCTGGAAGTGGAGCCCAGTGACACCATCGAAAATGTGAAGGCC
1574       ------+---------+---------+---------+---------+---------+---  1633
           TGGCCGTTCTGGTAGTGAGACCTTCACCTCGGGTCACTGTGGTAGCTTTTACACTTCCGG

a            R  Q  D  H  H  S  G  S  G  A  Q  *  H  H  R  K  C  E  G  Q  -
b          T  G  K  T  I  T  L  E  V  E  P  S  D  T  I  E  N  V  K  A    -
c            P  A  R  P  S  L  W  K  W  S  P  V  T  P  S  K  M  *  R  P  -

           AAGATCCAAGATAAAGAAGGCATCCCTCCCGACCAGCAGAGGCTCATCTTTGCAGGCAAG
1634       ------+---------+---------+---------+---------+---------+---  1693
           TTCTAGGTTCTATTTCTTCCGTAGGGAGGGCTGGTCGTCTCCGAGTAGAAACGTCCGTTC
```

```
a          D  P  R  *  R  R  H  P  S  R  P  A  E  A  H  L  C  R  Q  A  -
b        K  I  Q  D  K  E  G  I  P  P  D  Q  Q  R  L  I  F  A  G  K  -
c         R  S  K  I  K  K  A  S  L  P  T  S  R  G  S  S  L  Q  A  S  -

         CAGCTGGAAGATGGCCGCACTCTTTCTGACTACAACATCCAGAAGGAGTCGACCCTGCAC
    1694 ------+---------+---------+---------+---------+---------+--- 1753
         GTCGACCTTCTACCGGCGTGAGAAAGACTGATGTTGTAGGTCTTCCTCAGCTGGGACGTG

a          A  G  R  W  P  H  S  F  *  L  Q  H  P  E  G  V  D  P  A  P  -
b        Q  L  E  D  G  R  T  L  S  D  Y  N  I  Q  K  E  S  T  L  H  -
c         S  W  K  M  A  A  L  F  L  T  T  T  S  R  R  S  R  P  C  T  -

         CTGGTCCTGCGCCTGAGGGGTGGCTGTTAATTCTTCAGTCATGGCAT
    1754 ------+---------+---------+---------+---------+ 1800
         GACCAGGACGCGGACTCCCCACCGACAATTAAGAAGTCAGTACCGTA

a          G  P  A  P  E  G  W  L  L  I  L  Q  S  W  H  -
b        L  V  L  R  L  R  G  G  C  ⊛  F  F  S  H  G  -
c         W  S  C  A  *  G  V  A  V  N  S  S  V  M  A  -

Enzymes that do cut:

  NONE

Enzymes that do not cut:

    NotI
```

FIG. 4(III)

FIG. 5

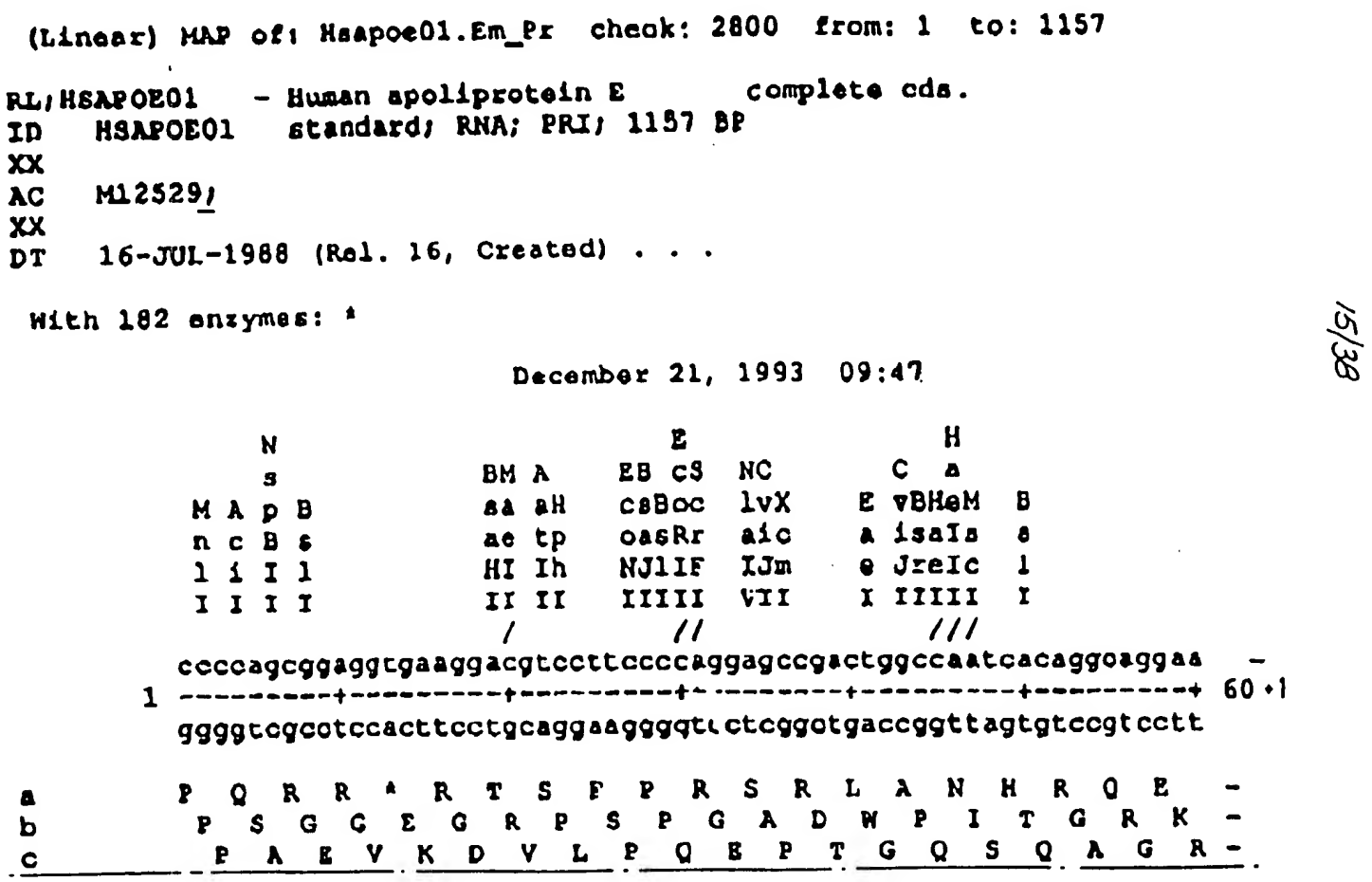

(Linear) MAP of: Hsapoe01.Em_Pr check: 2800 from: 1 to: 1157
RL;HSAPOE01 - Human apoliprotein E complete cds.
ID HSAPOE01 standard; RNA; PRI; 1157 BP
XX
AC M12529;
XX
DT 16-JUL-1988 (Rel. 16, Created) . . .
With 182 enzymes: *
December 21, 1993 09:47
N
s
M A p B
n c B s
l i I l
I I I I
BM A
aa aH
ae tp
HI Ih
II II
/
E
EB cS
csBoc
oasRr
NJlIF
IIIII
//
NC
lvX
aic
IJm
VII
H
C a
E vBHeM
a isaIs
e JreIc
I IIIII
///
B
s
l
I
ccccagcggaggtgaaggacgtccttccccaggagccgactggccaatcacaggcaggaa
1
ggggtcgcctccacttcctgcaggaaggggtctctcggctgaccggttagtgtccgtcctt
60
a P Q R R * R T S F P R S R L A N H R Q E
b P S G G E G R P S P G A D W P I T G R K
c P A E V K D V L P Q E P T G Q S Q A G R

FIG. 5(1)

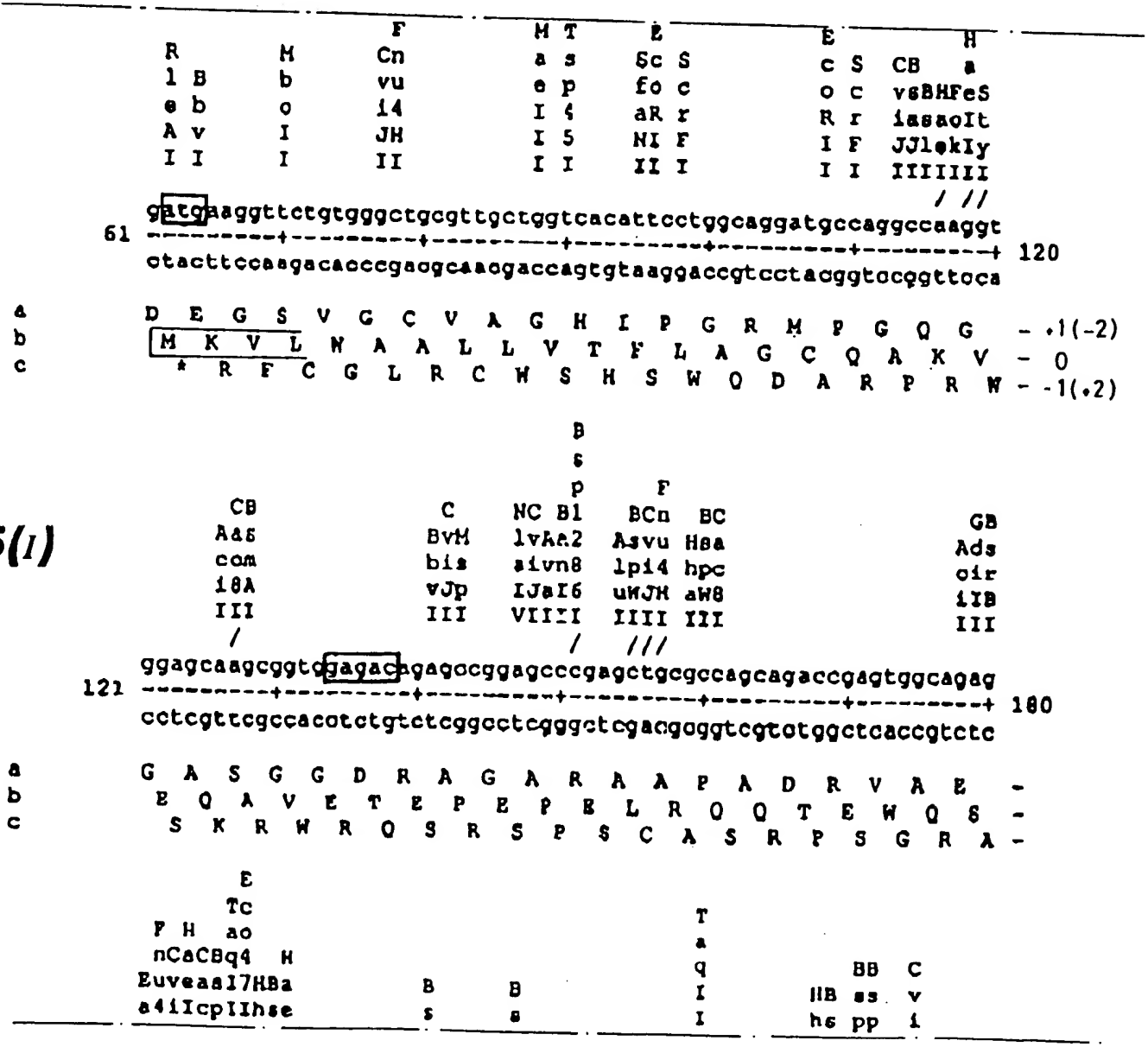


FIG. 5(II)

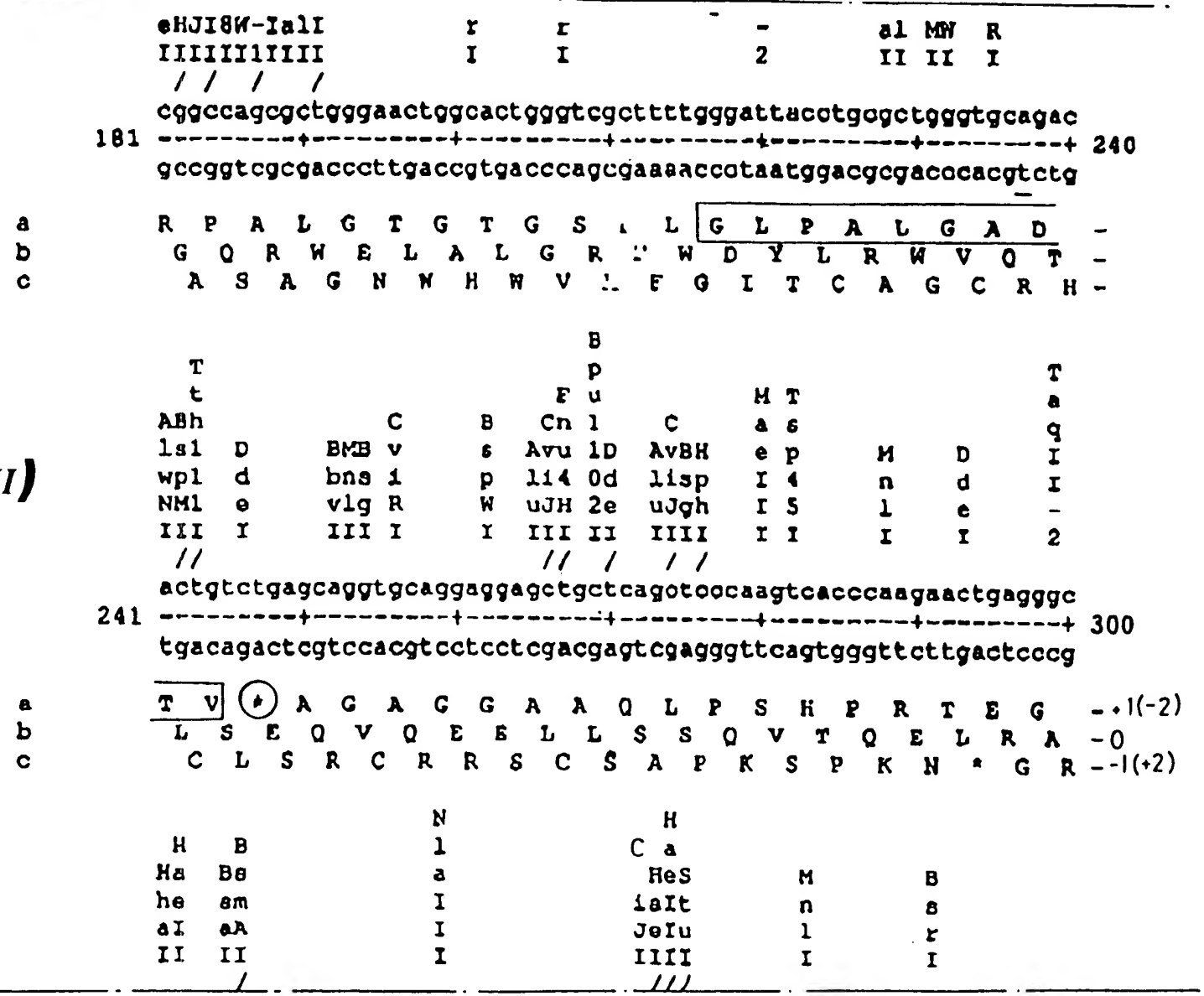
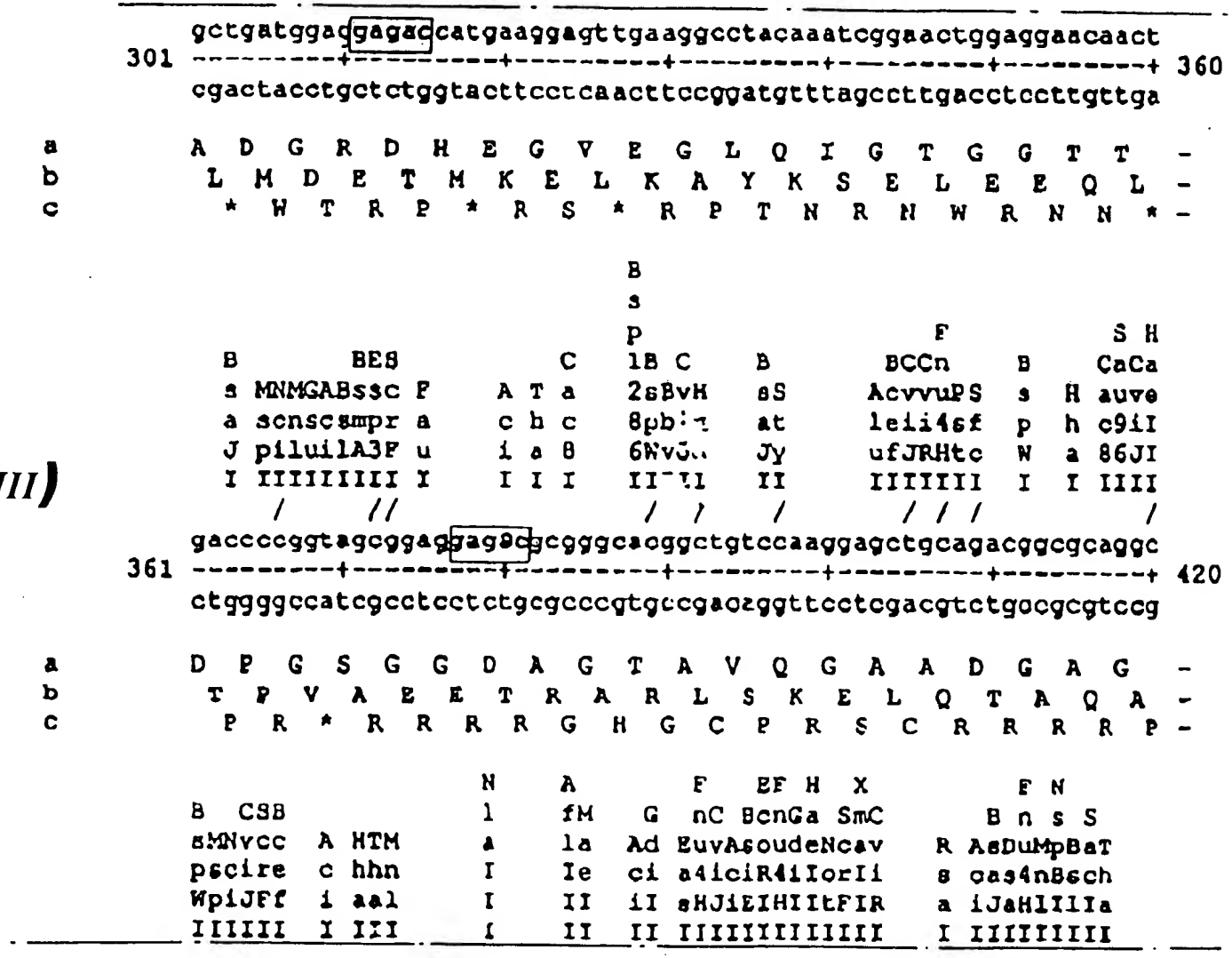


FIG. 5(III)


gctgatggaggagacgcatgaaggagttgaaggcctacaaatcggaactggaggaacaact
301
cgactacctgctctggtacttcctcaacttccggatgtttagccttgacctccttgttga
360
a A D G R D H E G V E G L Q I G T G G T T -
b L M D E T M K E L K A Y K S E L E E Q L -
c * W T R P * R S * R P T N R N W R N N * -
gaccccggtagcggaggagcgcgggcaaggctgtccaaggagctgcagacggcgcaggc
361
ctggggccatcgcctcctctgcgcccgtgccgaccggttcctcgacgtctgccgcgtccg
420
a D P G S G G D A G T A V Q G A A D G A G -
b T P V A E E T R A R L S K E L Q T A Q A -
c P R * R R R R G H G C P R S C R R R R P -

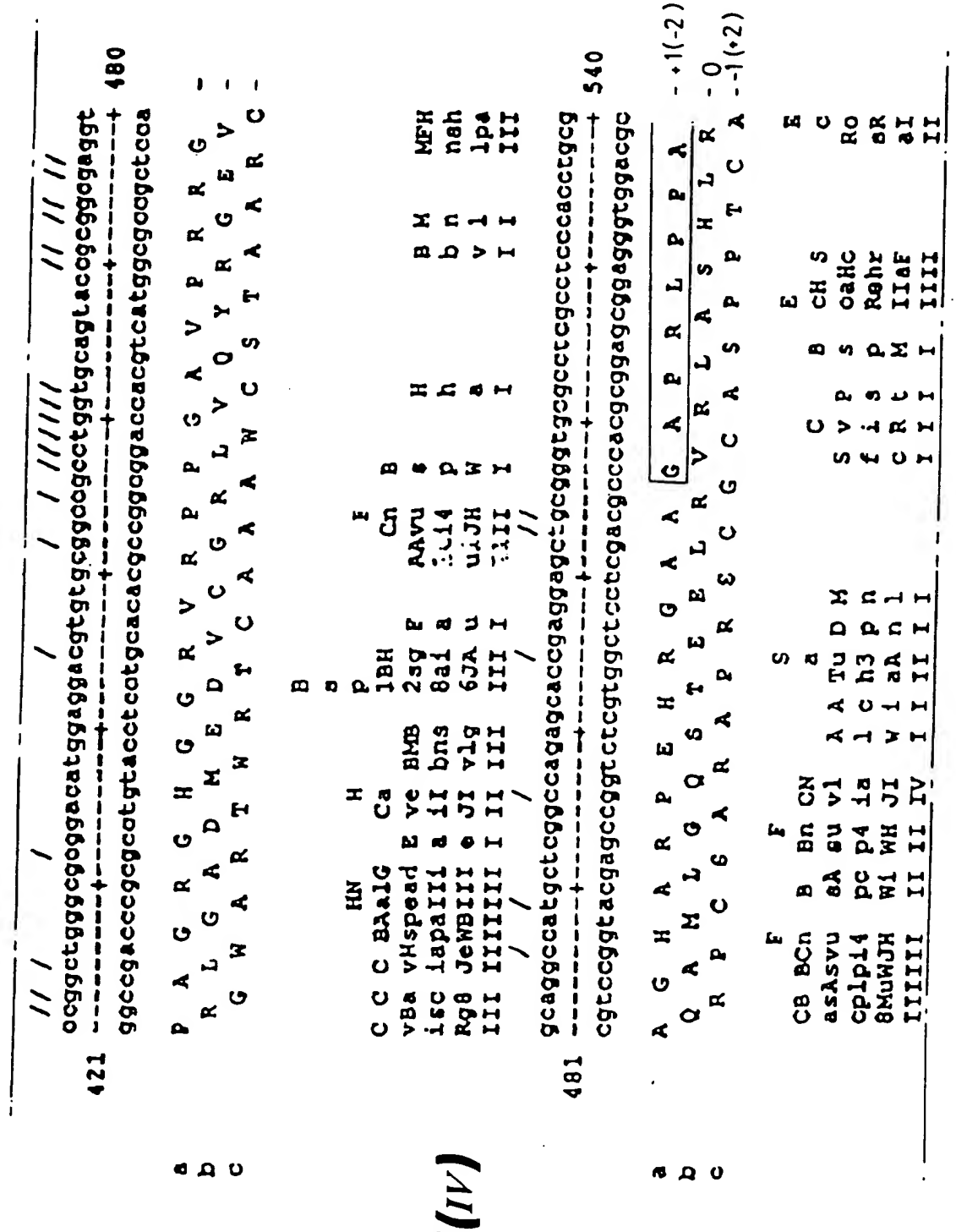


FIG. 5(IV)

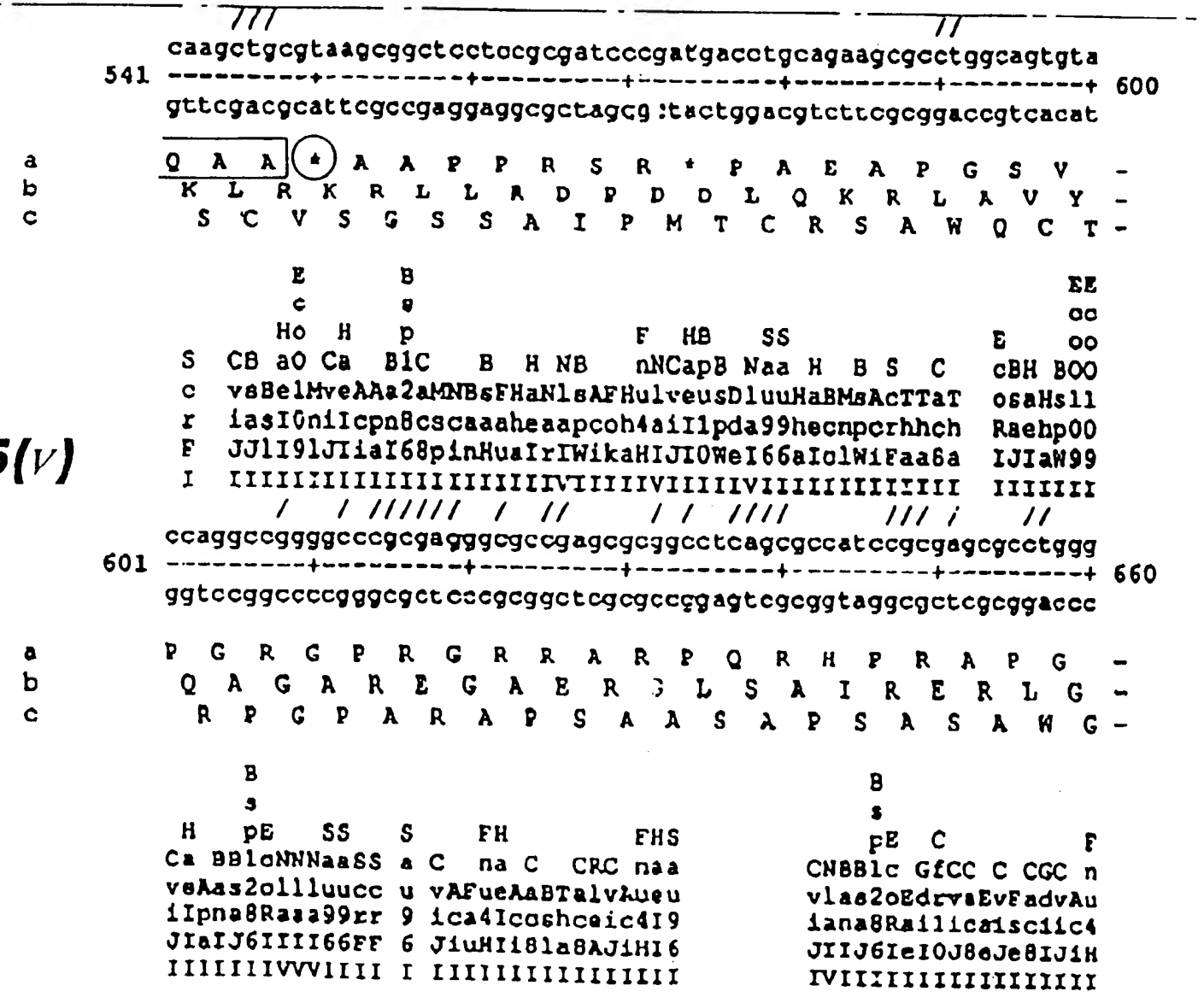


FIG. 5(V)

FIG. 6

```
ID   HSMAP2C   STANDARD;  RNA;  PRI;  5595BP
XX
AC   L12563;
XX
DT   20-NOV-1993 (REL. 37, CREATED)
DT   20-NOV-1993 (REL. 37,  LAST UPDATED, VERSION 1)
XX
DE   HOMO SAPIENS MICROTUBULE-ASSOCIATED PROTEIN ISOFORMS (MAP2 AND
DE   MAP2C) mRNA, COMPLETE CDS.
XX
KW   MICROTUBULE-ASSOCIATED PROTEIN
XX
OS   HOMO SAPIENS (HUMAN)
OC   EUKARYOTA; ANIMALIA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA
OC   THERIA; EUTHERIA; PRIMATES; HAPLORHINI; CATARRHINI; HOMINIDAE.
XX
RN   (1)
RP   1 - 5595
RA   ALBALA J.S., KALCHEVA N., SHAFIT-ZAGARDO B.;
RT   "CHARACTERIZATION OF THE TRANSCRIPT ENCODING 2 ISOFORMS OF HUMAN
RT   MICROTUBULE-ASSOCIATED PROTEIN 2 (MAP 2)";
RL   UNPUBLISHED
```

- MORE--7%)

FIG. 6(I)

```
                gcagggaagagtggtacctcaacacccactacccctgggtctactgccatcactcctggc
          4741  ---------+---------+---------+---------+---------+---------+ 4800
                cgtcccttctcaccatggagttgtgggtgatggggacccagatgacggtagtgaggaccg

a                AlaGlyLysSerGlyThrSerThrProThrThrProGlySerThrAlaIleThrProGly   -
b                 GlnGlyArgValValProGlnHisProLeuProLeuGlyLeuLeuProSerLeuLeuAla  -
c                  ArgGluGluTrpTyrLeuAsnThrHisTyrProTrpValTyrCysHisHisSerTrpHis -

                accccaccaagttattcttcacgcacaccaggcactcctggaacccctagctatcccagg
          4801  ---------+---------+---------+---------+---------+---------+ 4860
                tggggtggttcaataagaagtgcgtgtggtccgtgaggaccttggggatcgatagggtcc

a                ThrProProSerTyrSerSerArgThrProGlyThrProGlyThrProSerTyrProArg   -
b                 ProHisGlnValIleLeuHisAlaHisGlnAlaLeuLeuGluProLeuAlaIleProGly  -
c                  ProThrLysLeuPhePheThrHisThrArgHisSerTrpAsnProEndLeuSerGlnAsp -

                acccctcacacaccaggaaccccaagtctgccatcttggtgccgagtgagaagaaggtc
          4861  ---------+---------+---------+---------+---------+---------+ 4920
                tggggagtgtgtggtccttgggggttcagacggtagaaccacggctcactcttcttccag

a                ThrProHisThrProGlyThrProLysSerAlaIleLeuValProSerGluLysLysVal   -
b                 ProLeuThrHisGlnGluProProSerLeuProSerTrpCysArgValArgArgArgSer  -
c                  ProSerHisThrArgAsnProGlnValCysHisLeuGlyAlaGluEndGluGluGlyArg -

                gccatcatacgtactcctccaaaatctcctggactgactcccaagcagcttcggcttatt
          4921  ---------+---------+---------+---------+---------+---------+ 4980
                cggtagtatgcatgaggaggttttagaggacctgactgagggttcgtcgaagccgaataa

a                AlaIleIleArgThrProProLysSerProGlyLeuThrProLysGlnLeuArgLeuIle   -
b                 ProSerTyrValLeuLeuGlnAsnLeuLeuAspEndLeuProSerSerPheGlyLeuLeu  -
c                  HisHisThrTyrSerSerLysIleSerTrpThrAspSerGlnAlaAlaSerAlaTyrEnd -
```

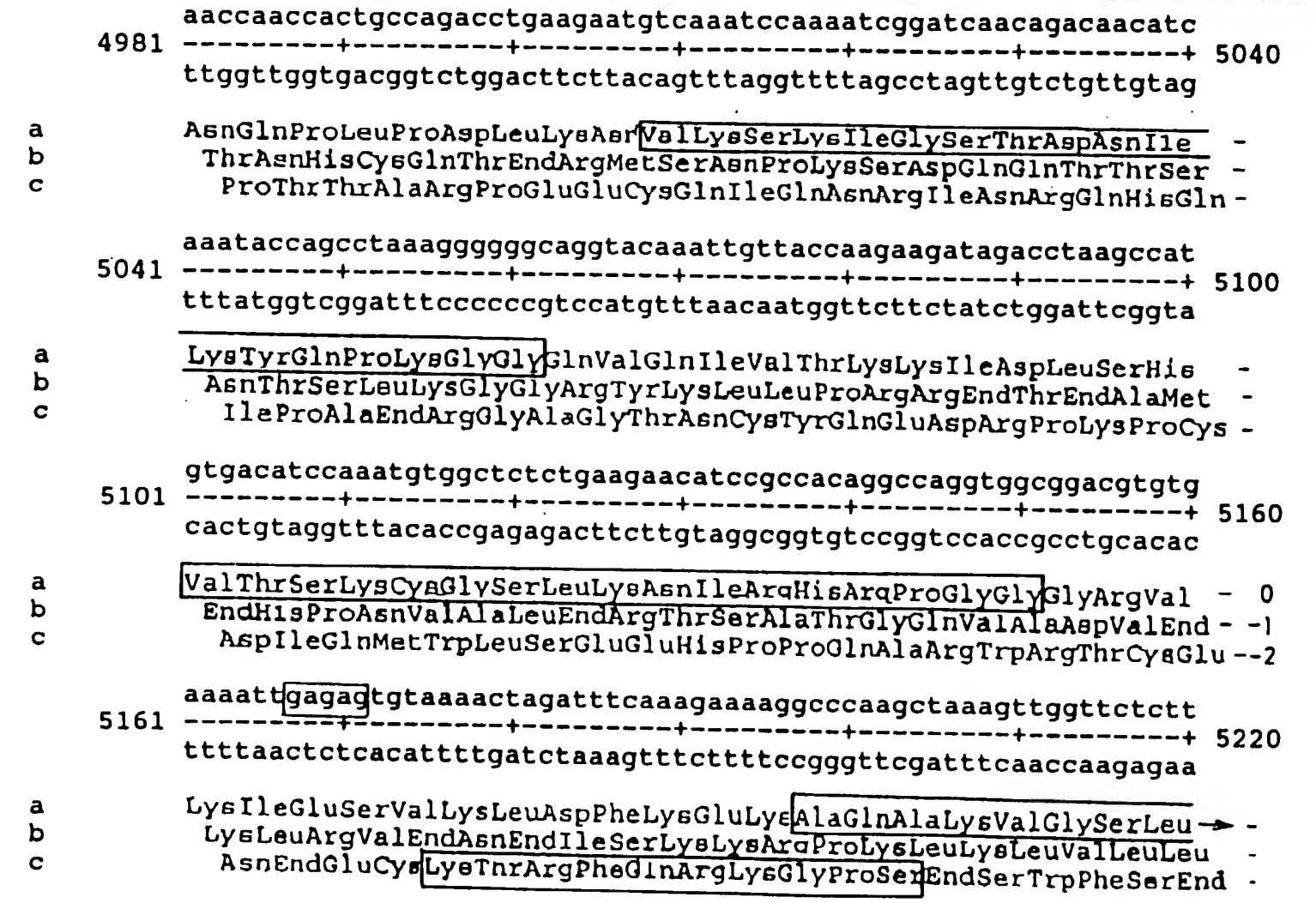


FIG. 6(II)

FIG. 7

```
(Linear) MAP of: Hsnflg.Em_Pr  check: 7027  from: 1  to: 4542

DL;HSNFLG      - Human gene for neurofilament subunit NF-LOW
ID   HSNFLG      standard; DNA; PRI; 4542 BP.
XX
AC   X05608;
XX
DT   29-NOV-1987 (Rel. 14, Created) . . .

With 1 enzymes: NOTI

                                November 29, 1993  18:39  ..

        gcagctcctcgggccgtagctcgacccgccttccctttccgcagaatcctcgccttgg
      1 ---------+---------+---------+---------+---------+---------+ 60
        cgtcgaggagcccggcatcgagctggggcggaagggaaaaggcgtcttaggagcggaacc

a       A  A  P  R  A  V  A  R  P  R  L  P  F  S  A  E  S  S  P  W  -
b        Q  L  L  G  P  *  L  D  P  A  F  P  F  P  Q  N  P  R  L  G  -
c         E  S  S  G  R  S  S  T  P  P  S  L  F  R  R  I  L  A  L  A -

        ctgcagcagcgcgctgcccccactggccggcgtgccgtgatcgatcgcaggctgcgtcag
     61 ---------+---------+---------+---------+---------+---------+ 120
        gacgtcgtcgcgcgacgggggtgaccggccgcacggcactagctagcgtccgacgcagtc

a       L  Q  Q  R  A  A  P  T  G  R  R  A  V  I  D  R  R  L  R  Q  -
b        C  S  S  A  L  P  P  L  A  G  V  P  *  S  I  A  G  C  V  R  -
c         A  A  A  R  C  P  H  W  P  A  C  R  D  R  S  Q  A  A  S  G -

        gacctcccggcgtataaatagggtggcagaacggcgccgagccgcacacagccatccat
    121 ---------+---------+---------+---------+---------+---------+ 180
        ctggagggccgcatatttatcccaccgtcttgccgcggctcggcgtgtgtcggtaggta

a       D  L  P  A  Y  K  *  G  W  Q  N  G  A  E  P  H  T  A  I  H  -
b        T  S  R  R  I  N  R  G  G  R  T  A  P  S  R  T  Q  P  S  I  -
c         P  P  G  V  *  I  G  V  A  E  R  R  R  A  A  H  S  H  P  S -
```

```
      cctccccttccctctctccoctgtcctctctctccgggctcccaccgccgccggggagc
181   ---------+---------+---------+---------+---------+---------+ 240
      ggaggggaagggagagaggggacaggagagagaggcccgagggtggcggcggcccctcg

a     P  P  P  S  L  S  P  L  S  S  L  S  G  L  P  P  P  P  G  S   -
b      L  P  L  P  S  L  P  C  P  L  S  P  G  S  H  R  R  R  G  A  -
c       S  P  F  P  L  S  P  V  L  S  L  R  A  P  T  A  A  G  E  H -

      accggccgccaaccaatgagttccttcagctacgagccgtactactcgacctcctacaag
241   ---------+---------+---------+---------+---------+---------+ 300
      tggccggcggttggttactcaaggaagtcgatgctcggcatgatgagctggaggatgttc

a     T  G  R  Q  P  M  S  S  F  S  Y  E  P  Y  Y  S  T  S  Y  K   -
b      P  A  A  N  Q  *  V  P  S  A  T  S  R  T  T  R  P  P  T  S  -
c       R  P  P  T  N  E  F  L  Q  L  R  A  V  L  L  D  L  L  Q  A -

      cggcgctacgtggagacgcccogggtgcatatcagcgtgcgcagcggctacagcaccgca
301   ---------+---------+---------+---------+---------+---------+ 360
      gccgcgatgcacctctgcgggggcccacgtatagtcgcacgcgtcgccgatgtcgtggcgt

a     R  R  Y  V  E  T  P  R  V  H  I  S  V  R  S  G  Y  S  T  A   -
b      G  A  T  W  R  R  P  G  C  I  S  A  C  A  A  A  T  A  P  H  -
c       A  L  R  G  D  A  P  G  A  Y  Q  R  A  Q  R  L  Q  H  R  T -

      cgctcagcttactcaagctactcggcgccggtgtcttcctcgctgtccgtgcgccgcagc
361   ---------+---------+---------+---------+---------+---------+ 420
      gcgagtcgaatgagttcgatgagccgcggccacagaaggagcgacaggcacgcggcgtcg

a     R  S  A  Y  S  S  Y  S  A  P  V  S  S  S  L  S  V  R  R  S   -
```

FIG. 7(I)

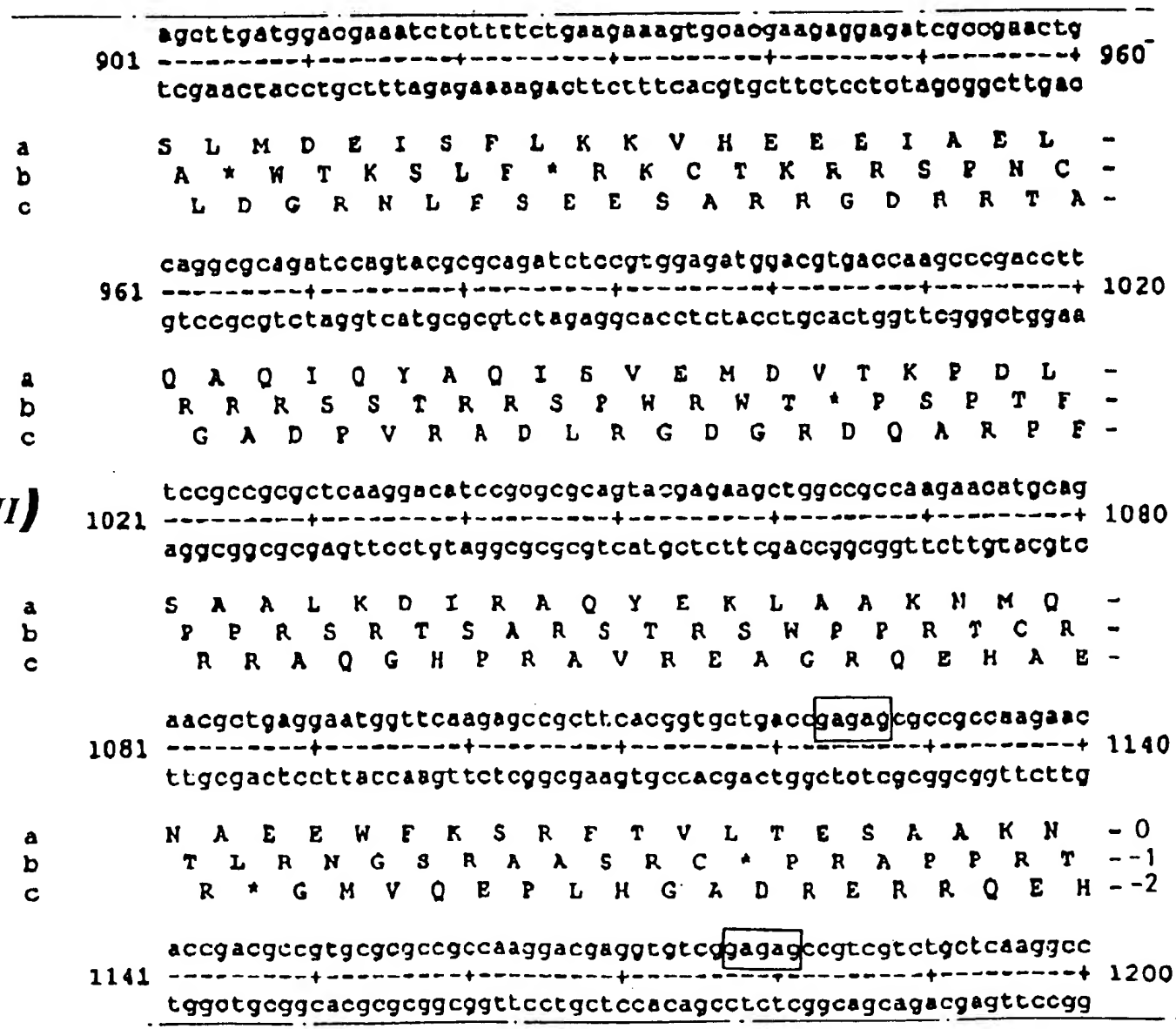


FIG. 7(II)

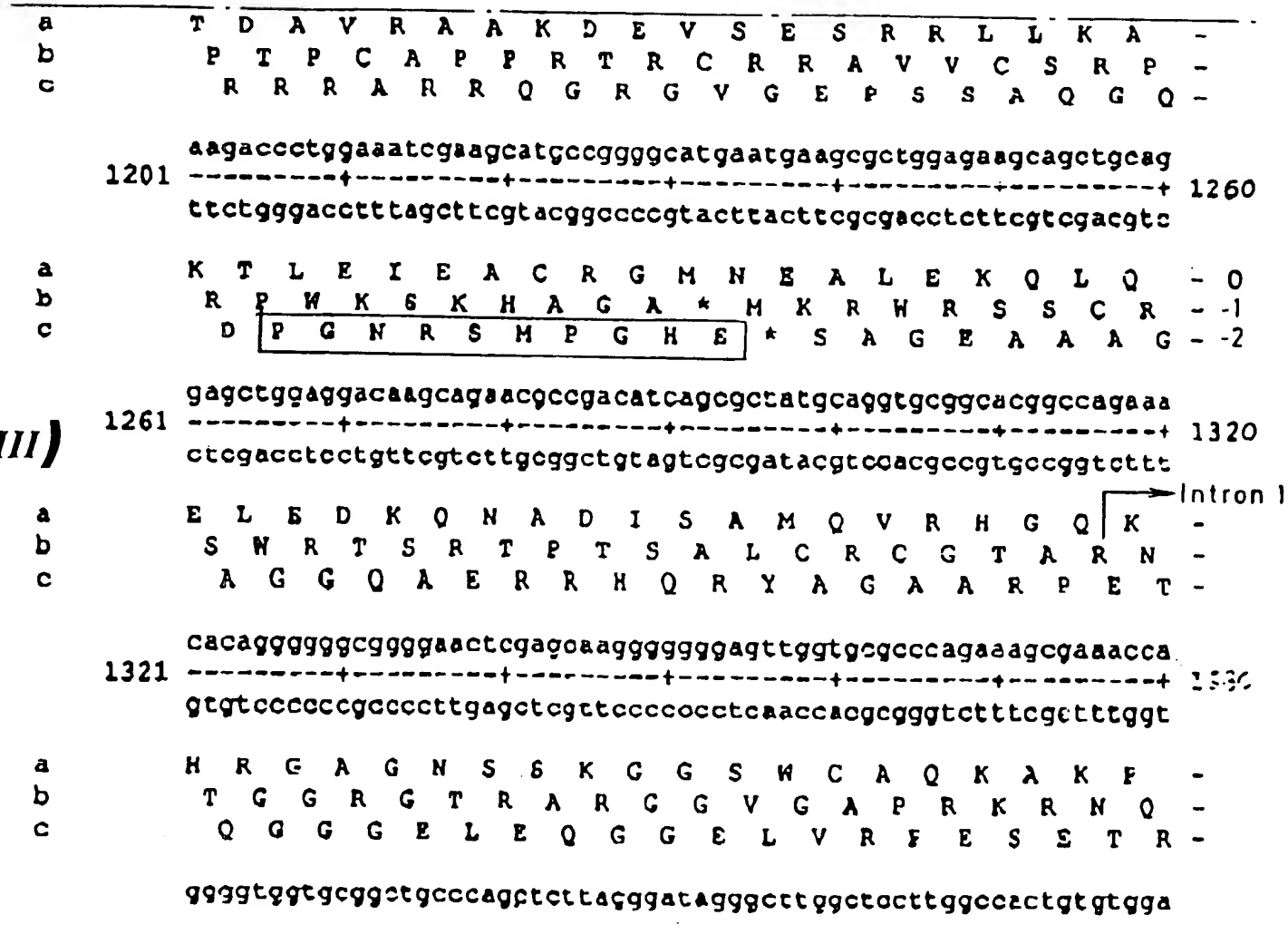

a T D A V R A A K D E V S E S R R L L K A -
b P T P C A P P R T R C R R A V V C S R P -
c R R R A R R Q G R G V G E P S S A Q G Q -
aagaccctggaaatcgaagcatgccgggcatgaatgaagcgctggagaagcagctgcag
1201
ttctgggacctttagcttcgtacggccccgtacttacttcgcgacctcttcgtcgacgtc
1260
a K T L E I E A C R G M N E A L E K Q L Q - 0
b R P W K S K H A G A * M K R W R S S C R - -1
c D P G N R S M P G H E * S A G E A A A G - -2
gagctggaggacaagcagaacgccgacatcagcgctatgcaggtgcggcacggccagaaa
1261
ctcgacctcctgttcgtcttgcggctgtagtcgcgatacgtccacgccgtgccggtcttt
1320
Intron 1
a E L E D K Q N A D I S A M Q V R H G Q K -
b S W R T S R T P T S A L C R C G T A R N -
c A G G Q A E R R H Q R Y A G A A R P E T -
cacaggggggcggggaactcgagcaaggggggggagttggtgcgcccagaaagcgaaacca
1321
gtgtccccccgccccttgagctcgttccccccctcaaccacgcgggtctttcgctttggt
a H R G A G N S S K G G S W C A Q K A K F -
b T G G R G T R A R G G V G A P R K R N Q -
c Q G G G E L E Q G G E L V R F E S E T R -
ggggtggtgcggctgcccagctcttacggatagggcttggctccttggccactgtgtgga

FIG. 7(III)

FIG. 8

(Linear) MAP of: Hsnfm.Em_Pr check: 3606 from: 1 to: 6236

DL;HSNFM - Human gene for neurofilament subunit M(NF-M)
ID HSNFM standard; DNA; PRI; 6236 BP.
XX
AC y0067;
XX
DT 19-SEP-1987 (Rel. 13, Created) . . .

With enzymes: NOTI

November 29, 1993 18.40 ..

```
          cagctgctttaagacaaggggtgggggaagggagggaggcaagaaaagatgagggtggg
        1 ---------+---------+---------+---------+---------+---------+ 60
          gtcgacgaaattctgttccccacccccttcccctccctccgttcttttctactcccaccc

a         Q  L  L  *  D  K  G  W  G  K  G  R  E  A  R  K  D  E  G  G  -
b          S  C  F  K  T  R  G  G  G  R  G  G  R  Q  E  K  M  R  V  G -
c           A  A  L  R  Q  G  V  G  E  G  E  G  G  K  K  R  *  G  W  G -

          ggaggggaaaagagggaatgcaaggggaaggagggaggagacggggagaaggaaagattg
       61 ---------+---------+---------+---------+---------+---------+ 120
          cctccccttttctcccttacgttcccccttcctccctcctctgcccctcttcctttctaac

a         G  G  E  K  R  E  C  K  G  K  E  G  G  D  G  E  K  E  R  L  -
b          E  G  K  R  G  N  A  R  G  R  R  E  E  T  G  R  R  K  D  W -
c           R  G  K  E  G  M  Q  G  E  G  G  R  R  R  G  E  G  K  I  G -

          gaagaaaaggatctccgaggaaggggctgagagaagggcagggtgaactggactaaaggc
      121 ---------+---------+---------+---------+---------+---------+ 180
          cttcttttcctagaggctccttccccgactctcttcccgtcccacttgacctgatttccg
```

FIG. 8(I)

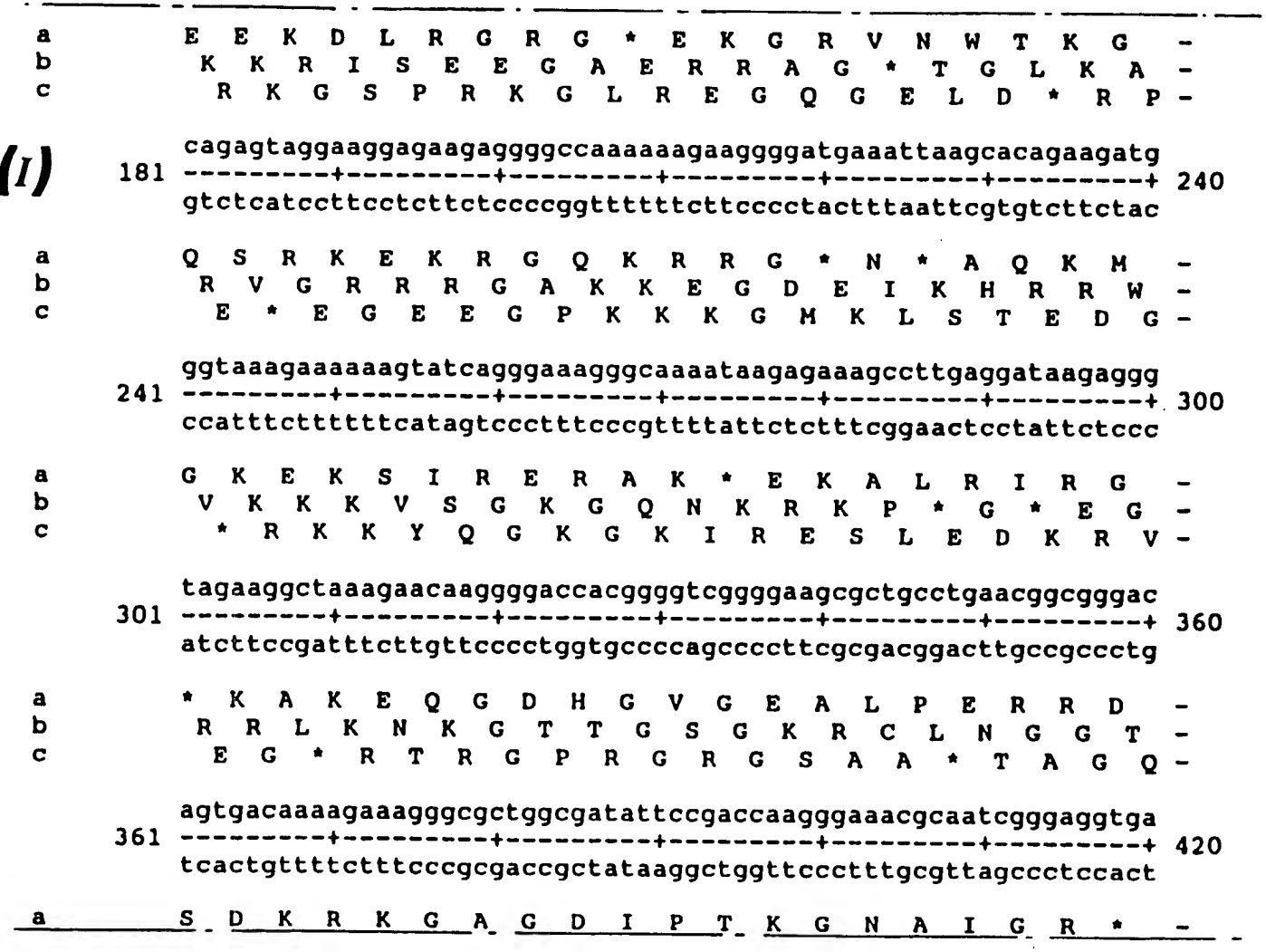

a E E K D L R G R G * E K G R V N W T K G -
b K K R I S E E G A E R R A G * T G L K A -
c R K G S P R K G L R E G Q G E L D * R P -
cagagtaggaaggagaagaggggccaaaaaagaaggggatgaaattaagcacagaagatg
181
240
gtctcatccttcctcttctccccggtttttttcttcccctactttaattcgtgtcttctac
a Q S R K E K R G Q K R R G * N * A Q K M -
b R V G R R R G A K K E G D E I K H R R W -
c E * E G E E G P K K K G M K L S T E D G -
ggtaaagaaaaaagtatcagggaaagggcaaaataagagaaagccttgaggataagaggg
241
300
ccatttctttttcatagtccctttcccgttttattctctttcggaactcctattctccc
a G K E K S I R E R A K * E K A L R I R G -
b V K K K V S G K G Q N K R K P * G * E G -
c * R K K Y Q G K G K I R E S L E D K R V -
tagaaggctaaagaacaaggggaccacggggtcggggaagcgctgcctgaacggcgggac
301
360
atcttccgatttcttgttcccctggtgccccagccccttcgcgacggacttgccgccctg
a * K A K E Q G D H G V G E A L P E R R D -
b R R L K N K G T T G S G K R C L N G G T -
c E G * R T R G P R G R G S A A * T A G Q -
agtgacaaaagaaagggcgctggcgatattccgaccaagggaaacgcaatcgggaggtga
361
420
tcactgttttctttcccgcgaccgctataaggctggttccctttgcgttagccctccact
a S D K R K G A G D I P T K G N A I G R * -

FIG. 8(II)

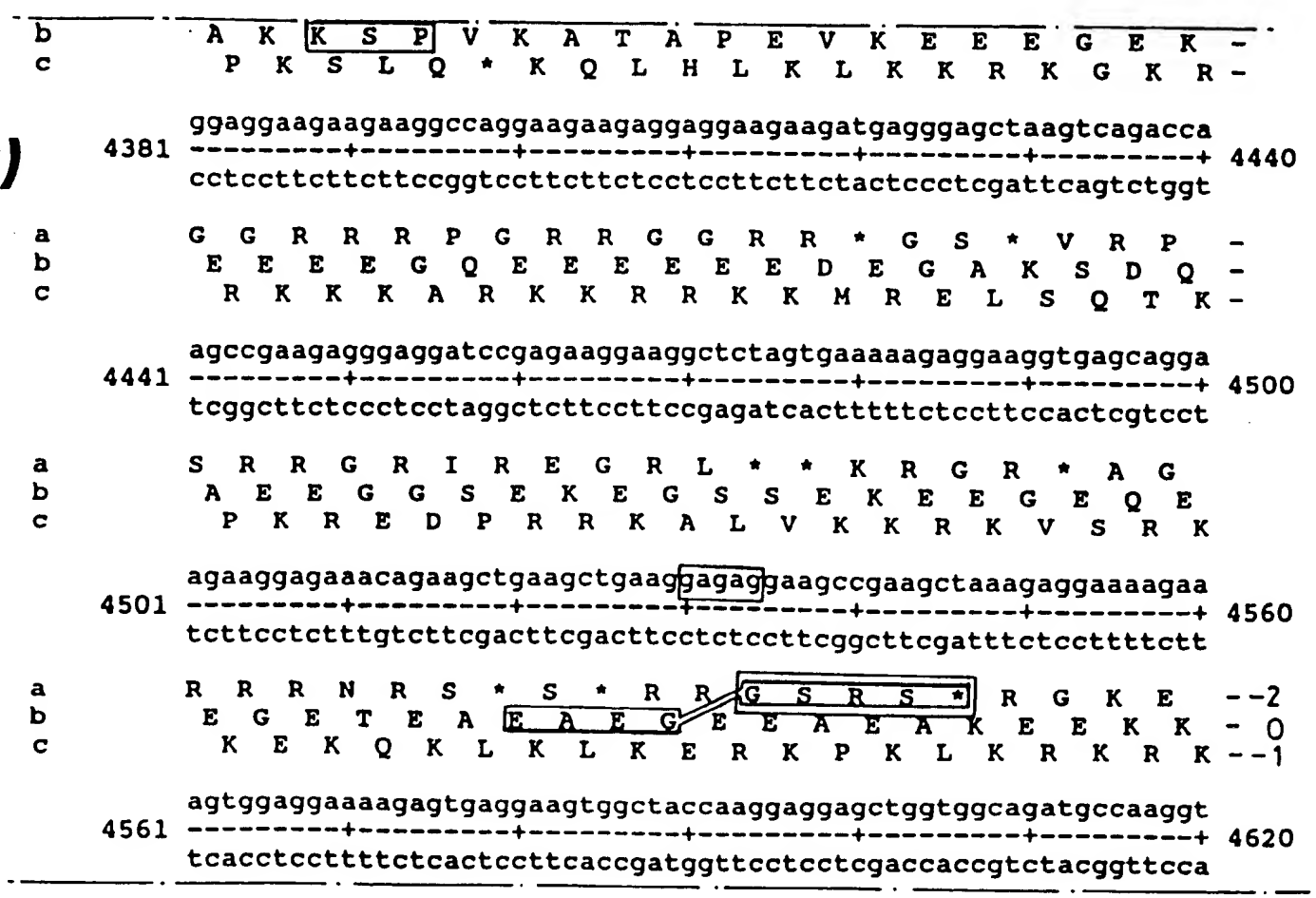


FIG. 8(III)

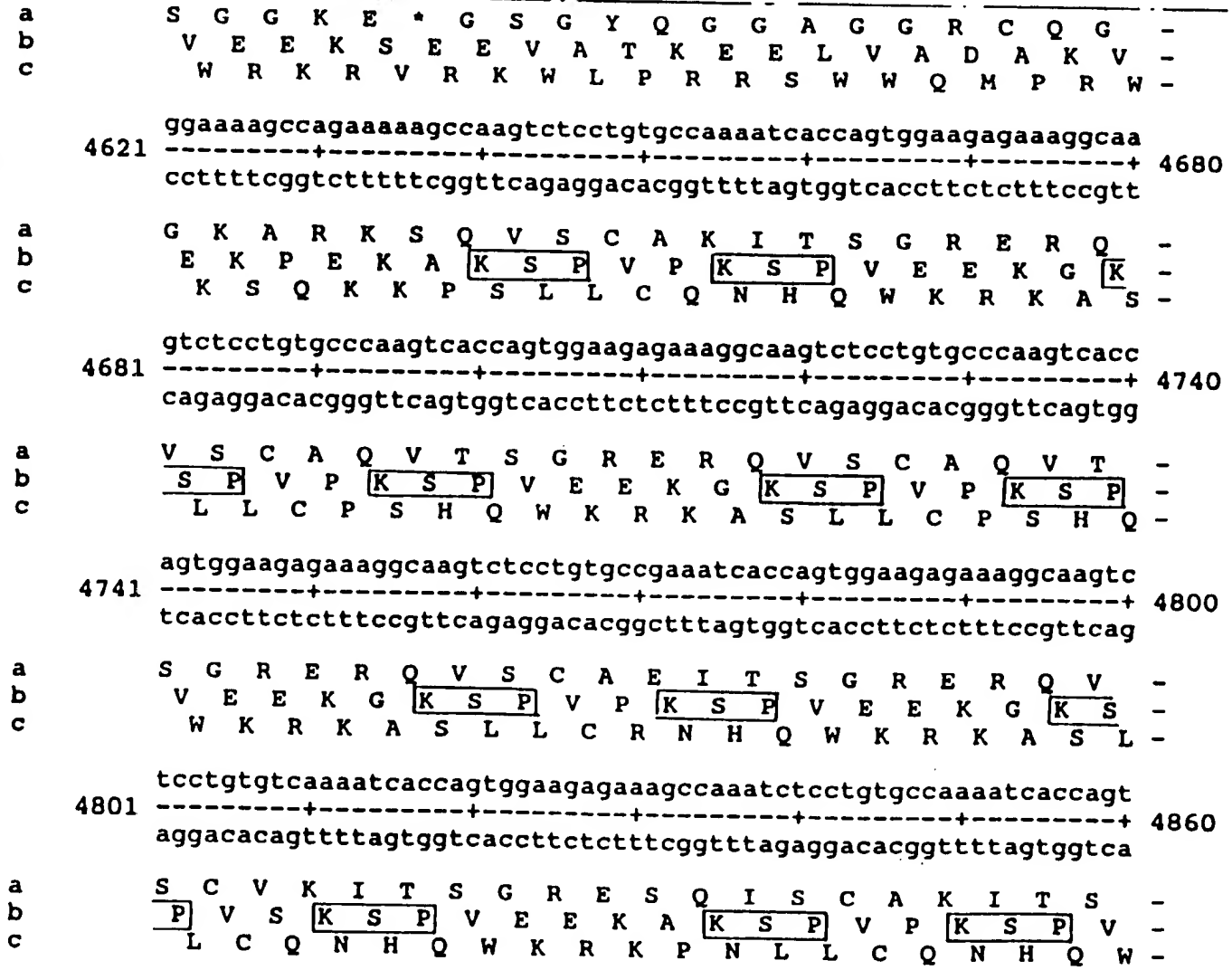

a S G G K E * G S G Y Q G G A G G R C Q G -
b V E E K S E E V A T K E E L V A D A K V -
c W R K R V R K W L P R R S W W Q M P R W -
4621 ggaaaagccagaaaaagccaagtctcctgtgccaaaatcaccagtggaagagaaaggcaa
cctttcggtcttttcggttcagaggacacggttttagtggtcaccttctctttccgtt 4680
a G K A R K S Q V S C A K I T S G R E R Q -
b E K P E K A K S P V P K S P V E E K G K -
c K S Q K K P S L L C Q N H Q W K R K A S -
4681 gtctcctgtgcccaagtcaccagtggaagagaaaggcaagtctcctgtgcccaagtcacc
cagaggacacgggttcagtggtcaccttctctttccgttcagaggacacgggttcagtgg 4740
a V S C A Q V T S G R E R Q V S C A Q V T -
b S P V P K S P V E E K G K S P V P K S P -
c L L C P S H Q W K R K A S L L C P S H Q -
4741 agtggaagagaaaggcaagtctcctgtgccgaaatcaccagtggaagagaaaggcaagtc
tcaccttctctttccgttcagaggacacggctttagtggtcaccttctctttccgttcag 4800
a S G R E R Q V S C A E I T S G R E R Q V -
b V E E K G K S P V P K S P V E E K G K S -
c W K R K A S L L C R N H Q W K R K A S L -
4801 tcctgtgtcaaaatcaccagtggaagagaaagccaaatctcctgtgccaaaatcaccagt
aggacacagttttagtggtcaccttctctttcggtttagaggacacggttttagtggtca 4860
a S C V K I T S G R E S Q I S C A K I T S -
b P V S K S P V E E K A K S P V P K S P V -
c L C Q N H Q W K R K P N L L C Q N H Q W -

FIG. 9

(Linear) MAP of: Humnfhl check: 1349 from: 1 to: 1162

```
LOCUS          HUMNFH1     1162bp ds-DNA              PRI       15-MAR-1990
DEFINITION     Human gene for heavy neurofilament subunit (NF-H) exon 1.
ACCESSION      X15306 X12501
KEYWORDS       intermediate filament; neurofilament; phosphoprotein.
SEGMENT        1 of 4
SOURCE         human (Homo sapiens) . . . .
```

With 1 enzymes: NOTI

January 6, 1994 16.05 ..

```
          CCACTCCGGAGTCCTCTGCCCGCTTCCCGACCTCGAGGGTCTCCTCTGACGCGCAGCGTC
        1 ---------+---------+---------+---------+---------+---------+ 60
          GGTGAGGCCTCAGGAGACGGGCGAAGGGCTGGAGCTCCCAGAGGAGACTGCGCGTCGCAG

a           P  L  R  S  P  L  P  A  S  R  P  R  G  S  P  L  T  R  S  V  -
b            H  S  G  V  L  C  P  L  P  D  L  E  G  L  L  *  R  A  A  S  -
c             T  P  E  S  S  A  R  F  P  T  S  R  V  S  S  D  A  Q  R  R -

          GATTCCCCTTCCCTCCTCGGTCCCCTGCCCCGCCCCTCTCACTGCGCGGAGCCGGTCGCC
       61 ---------+---------+---------+---------+---------+---------+ 120
          CTAAGGGGAAGGGAGGAGCCAGGGGACGGGGCGGGGAGAGTGACGCGCCTCGGCCAGCGG

a           D  S  P  S  L  L  G  P  L  P  R  P  S  H  C  A  E  P  V  A  -
b            I  P  L  P  S  S  V  P  C  P  A  P  L  T  A  R  S  R  S  P  -
c             F  P  F  P  P  R  S  P  A  P  P  L  S  L  R  G  A  G  R  R -

          GGGGGGCCGCAGGGGAGGAGGCGGAGAGGCGGGGCCCTCCTCCCCACCCTCTCACTGCCA
      121 ---------+---------+---------+---------+---------+---------+ 180
          CCCCCCGGCGTCCCCTCCTCCGCCTCTCCGCCCCGGGAGGAGGGGTGGGAGAGTGACGGT
```

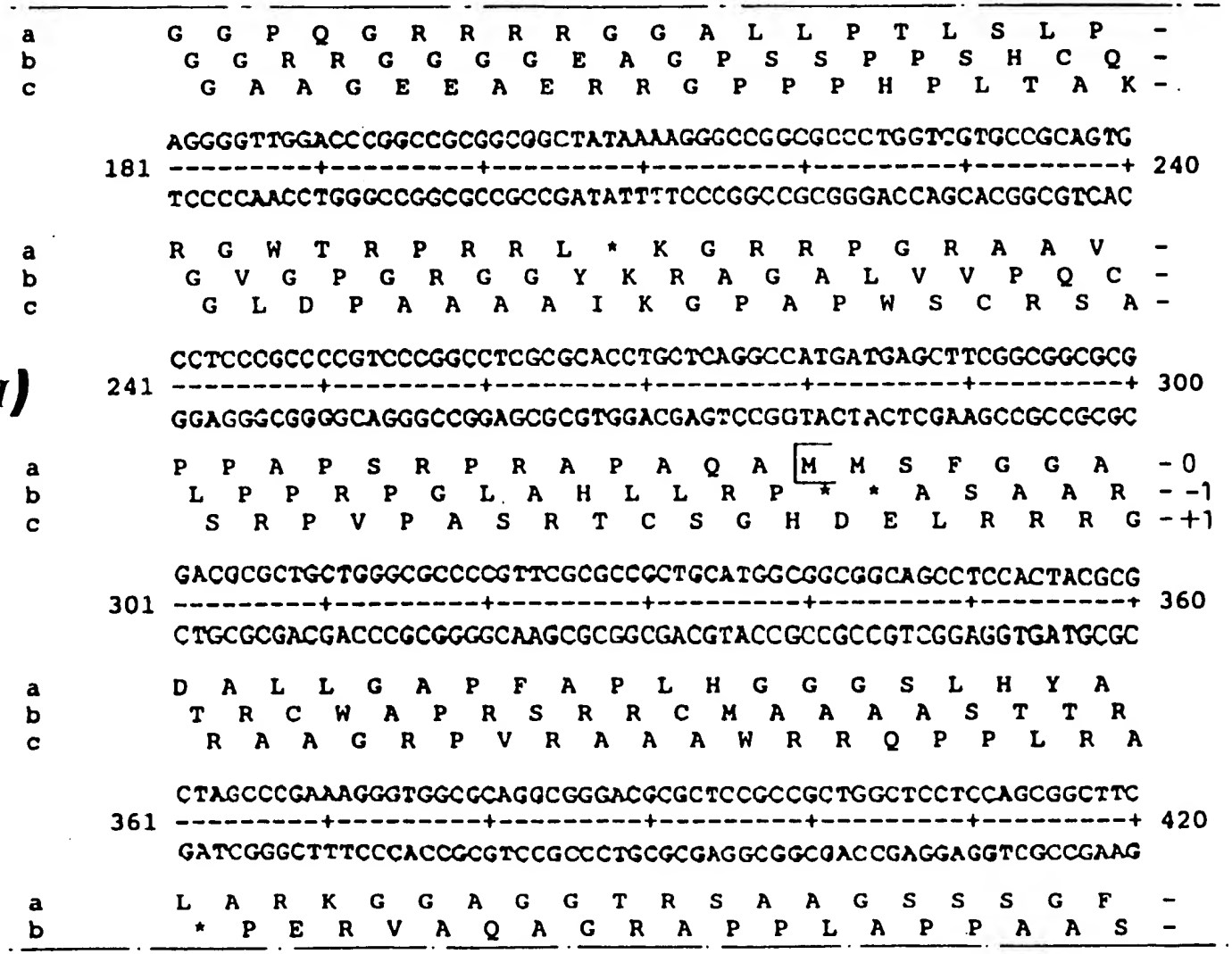


FIG. 9(I)

FIG. 9(II)

```
c          S  P  K  G  W  R  R  R  D  A  L  R  R  W  L  L  Q  R  L  P -

           CACTCGTGGACACGGACGTCCGTGAGCTCCGTGTCCGCCTCGCCCAGCCGCTTCCGTGGC
       421 ---------+---------+---------+---------+---------+---------+ 480
           GTGAGCACCTGTGCCTGCAGGCACTCGAGGCACAGGCGGAGCGGGTCGGCGAAGGCACCG

a          H  S  W  T  R  T  S  V  S  S  V  S  A  S  P  S  R  F  R  G  -
b           T  R  G  H  G  R  P  *  A  P  C  P  P  R  P  A  A  S  V  A  -
c           L  V  D  T  D  V  R  E  L  R  V  R  L  A  Q  P  L  P  W  R -

           GCAGGCGCCGCCTCAAGCACCGACTCGCTGGACACGCTGAGCAACGGGCCGGAGGGCTGC
       481 ---------+---------+---------+---------+---------+---------+ 540
           CGTCCGCGGCGGAGTTCGTGGCTGAGCGACCTGTGCGACTCGTTGCCCGGCCTCCCGACG

a          A  G  A  A  S  S  T  D  S  L  D  T  L  S  N  G  P  E  G  C  -
b           Q  A  P  P  Q  A  P  T  R  W  T  R  *  A  T  G  R  R  A  A  -
c           R  R  R  L  K  H  R  L  A  G  H  A  E  Q  R  A  G  G  L  H -

           ATGGTGGCGGTGGCCACCTCACGCAGTGAGAAGGAGCAGCTGCAGGCGCTGAACGACCGC
       541 ---------+---------+---------+---------+---------+---------+ 600
           TACCACCGCCACCGGTGGAGTGCGTCACTCTTCCTCGTCGACGTCCGCGACTTGCTGGCG

a          M  V  A  V  A  T  S  R  S  E  K  E  Q  L  Q  A  L  N  D  R  -
b           W  W  R  W  P  P  H  A  V  R  R  S  S  C  R  R  *  T  T  A  -
c           G  G  G  G  H  L  T  Q  *  E  G  A  A  A  G  A  E  R  P  L -

           TTCGCCGGGTACATCGACAAGGTGCGGCAGCTGGAGGCGCACAACCGCAGCCTGGAGGGC
       601 ---------+---------+---------+---------+---------+---------+ 660
           AAGCGGCCCATGTAGCTGTTCCACGCCGTCGACCTCCGCGTGTTGGCGTCGGACCTCCCG

a          F  A  G  Y  I  D  K  V  R  Q  L  E  A  H  N  R  S  L  E  G  -
b           S  P  G  T  S  T  R  C  G  S  W  R  R  T  T  A  A  W  R  A  -
c           R  R  V  H  R  Q  G  A  A  A  G  G  A  Q  P  Q  P  G  G  R -

           GAGGCTGCGGCGCTGCGGCAGCAGCAGGCGGGCCGCTCCGCTATGGGCGAGCTGTACGAG
       661 ---------+---------+---------+---------+---------+---------+ 720
           CTCCGACGCCGCGACGCCGTCGTCGTCCGCCCGGCGAGGCGATACCCGCTCGACATGCTC
```

```
a        E  A  A  A  L  R  Q  Q  Q  A  G  R  S  A  M  G  E  L  Y  E  -
b         R  L  R  R  C  G  S  S  R  R  A  A  P  L  W  A  S  C  T  S  -
c          G  C  G  A  A  A  A  G  G  P  L  R  Y  G  R  A  V  R  A  -

         CGCGAGGTCCGCGAGATGCGCGGCGCGGTGCTGCGCCTGGCGCGGCGCGCGGTCAGCTA
    721  ---------+---------+---------+---------+---------+---------+ 780
         GCGCTCCAGGCGCTCTACGCGCCGCGCCACGACGCGGACCCGCGCCGCGCGCCAGTCGAT

a        R  E  V  R  E  M  R  G  A  V  L  R  L  G  A  A  R  G  Q  L  -
b         A  R  S  A  R  C  A  A  R  C  C  A  W  A  R  R  A  V  S  Y  -
c          R  G  P  R  D  A  R  R  G  A  A  P  G  R  G  A  R  S  A  T  -
                    ->NF-H                <-
         CGCCTGGAGCAGGAGCACCTGCTCGAGGACATCGCGCACGTGCGCCAGCGCCTAGACGAC
    781  ---------+---------+---------+---------+---------+---------+ 840
         GCGGACCTCGTCCTCGTGGACGAGCTCCTGTAGCGCGTGCACGCGGTCGCGGATCTGCTG

a        R  L  E  Q  E  H  L  L  E  D  I  A  H  V  R  Q  R  L  D  D  -
b         A  W  S  R  S  T  C  S  R  T  S  R  T  C  A  S  A  *  T  T  -
c          P  G  A  G  A  P  A  R  G  H  R  A  R  A  P  A  P  R  R  R  -

         GAGGCCCGGCAGCGAGAGGAGGCCGAGGCGGCGGCCCGCGCGCTGGCGCGCTTCGCGCAG
    841  ---------+---------+---------+---------+---------+---------+ 900
         CTCCGGGCCGTCGCTCTCCTCCGGCTCCGCCGCCGGGCGCGCGACCGCGCGAAGCGCGTC

a        E  A  R  Q  R  E  E  A  E  A  A  A  R  A  L  A  R  F  A  Q  -0
b         R  P  G  S  E  R  R  P  R  R  R  P  A  R  W  R  A  S  R  R  --1
c          G  P  A  A  R  G  G  R  G  G  G  P  R  A  G  A  L  R  A  G  --2

         GAGGCCGAGGCGGCGCGCGTGGACCTGCAGAAGAAGGCGCAGGCGCTGCAGGAGGAGTGC
    901  ---------+---------+---------+---------+---------+---------+ 960
         CTCCGGCTCCGCCGCGCGCACCTGGACGTCTTCTTCCGCGTCCGCGACGTCCTCCTCACG
```

FIG. 9(III)

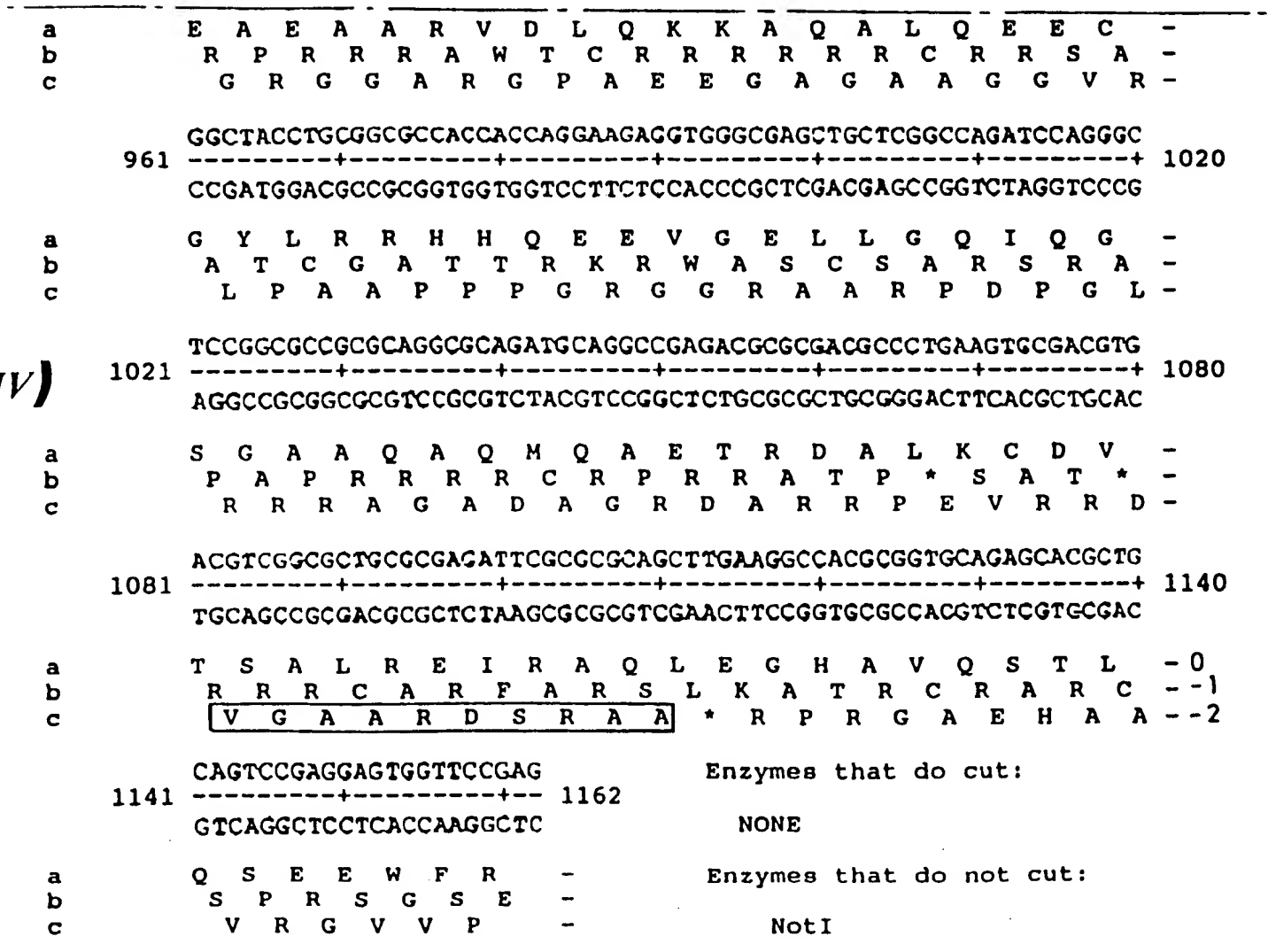

a E A E A A R V D L Q K K A Q A L Q E E C -
b R P R R R A W T C R R R R R R C R R S A -
c G R G G A R G P A E E G A G A A G G V R -
GGCTACCTGCGGCGCCACCACCAGGAAGAGGTGGGCGAGCTGCTCGGCCAGATCCAGGGC
961 + + + + + 1020
CCGATGGACGCCGCGGTGGTGGTCCTTCTCCACCCGCTCGACGAGCCGGTCTAGGTCCCG
a G Y L R R H H Q E E V G E L L G Q I Q G -
b A T C G A T T R K R W A S C S A R S R A -
c L P A A P P P G R G G R A A R P D P G L -
TCCGGCGCCGCGCAGGCGCAGATGCAGGCCGAGACGCGCGACGCCCTGAAGTGCGACGTG
1021 + + + + + 1080
AGGCCGCGGCGCGTCCGCGTCTACGTCCGGCTCTGCGCGCTGCGGGACTTCACGCTGCAC
a S G A A Q A Q M Q A E T R D A L K C D V -
b P A P R R R R C R P R R A T P * S A T * -
c R R R A G A D A G R D A R R P E V R R D -
ACGTCGGCGCTGCGCGACATTCGCGCGCAGCTTGAAGGCCACGCGGTGCAGAGCACGCTG
1081 + + + + + 1140
TGCAGCCGCGACGCGCTCTAAGCGCGCGTCGAACTTCCGGTGCGCCACGTCTCGTGCGAC
a T S A L R E I R A Q L E G H A V Q S T L -0
b R R R C A R F A R S L K A T R C R A R C --1
c V G A A R D S R A A * R P R G A E H A A --2
CAGTCCGAGGAGTGGTTCCGAG
1141 + + -- 1162
GTCAGGCTCCTCACCAAGGCTC
a Q S E E W F R -
b S P R S G S E -
c V R G V V P -
Enzymes that do cut:
NONE
Enzymes that do not cut:
NotI

FIG. 9(IV)

FIG. 10

β amyloid precursor protein (exons 9 and 10):

GAGAGGCTTGAGGCCAAGCACCGAGAGAGAATGTCCCAGGTCATGAGAGAATGGGAAGAGGCAGAACGTCAAGCAAAGAACTTGCCTAAA

E R L E A K H R E R M S Q V M R E W E E A E R Q A K N L P K wt

E A * G Q A P R E N V P G H E R M G R G R T S S K E L A * +1

Tau (exon 13):

GGAGATCGTGTACAAGTCGCCAGTGGTGTCTGGGGACACGTCTCCACGGCATCTCAGC

G D R V Q V A S G V W G H V S T A S Q Q

E I V Y K S P V V S G D T S P R H L S

AATGTCTCCTCCACCGGCAGCATCGACATGGTAGACTCGCCCCAGCTCGCCACGCTAGCTGAC

N V S S T G S I D M V D S P Q L A T L A D wt

C L L H R Q H R H G R L A P A R H A S * +1

Ubiquitin B (exon 2):

TCTGAGAGGTGGTATGCAGATCTTCGTGAAGACCCTGACCGGCAAGACCATCACCCTGGAAGTGGAGCCCAGTGA

L R L R G G M Q V K T L T G K T I T L E V E P S D wt

E R W Y A D L R E D P D R Q
R Q Q D H H P G S G A Q * +1

ubi B 1 —|— ubi B 2

Ubi B 2: antibody used for tables. Ubi B1 also gives rise to staining of tangles. However, its titer is much lower than that of Ubi B 2.

FIG. 10(I)